

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



**The relationship of theory of mind to executive function and language in children with autism and children with moderate learning difficulties**

Fisher, Naomi Catherine

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

**END USER LICENCE AGREEMENT**



**Unless another licence is stated on the immediately following page** this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

**Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

**The Relationship of Theory of Mind to Language and  
Executive Function in Children with Autism  
and Children with Moderate Learning Difficulties**

**Naomi Catharine Fisher**

Thesis submitted to the University of London for the  
degree of Doctor of Philosophy  
2002

Institute of Psychiatry  
King's College  
University of London

Prepared under the supervision of  
Francesca Happé and Judy Dunn



## **ACKNOWLEDGEMENTS**

I would like to express my gratitude to all the children and staff of the many schools who willingly gave up their time to take part in this research; and from whom I learnt so much. Without them this work would not have been possible. I would particularly like to thank Liam, who did not participate in this research, but who constantly and patiently teaches me something of what it is like to live with autism.

I would like to thank Dr. Francesca Happé for her support, advice and generosity with her time and ideas throughout this PhD. I would also like to thank Prof. Judy Dunn, for all her constructive criticism, interest and expertise.

Thank you to everyone who helped create a stimulating and supportive working environment, especially Essi, Greg, Rebecca, Rhonda and John, all of whose help was invaluable. Many thanks to Dr. Elisabeth Hill for her useful comments and for her encouragement.

Thanks to my family for all their confidence in me, and to all my friends, particularly Rebekah (who also proof-read, an incredible feat), Katharine, Philippa, Katy, Helen and Jessica.

## ABSTRACT

This thesis aimed to investigate the nature of theory of mind (ToM) and its relationship to language and executive function (EF) in children with autistic spectrum disorders (ASD) and children with moderate learning difficulties (MLD). A total of 188 children (70 with ASD, 118 with MLD) were tested over five studies. Study 1 looked at the relationship between language and ToM tasks, whilst study 2 investigated the relationship between EF and ToM in those who failed false belief tasks. Study 3 was a training study of EF and ToM, looking for direct and indirect effects of intervention. Study 4 investigated the performance on advanced tests of ToM of children with ASD who passed false belief tasks, and finally study 5 looked at the relationship between teacher ratings of real life behaviour of ToM and EF and performance on experimental tasks.

Impairments in ToM were more common and pervasive in children with ASD as compared to those with MLD. However, some individuals with ASD did pass ToM tasks, and there was evidence that these children had a superior understanding of mental states across a range of tasks and in real life. There was also some suggestion that failing a ToM task for a child with MLD may not be entirely spurious. Language, in particular grammar, was strongly related to ToM task performance in ASD, suggesting that it may be important developmentally as well as online. EF was found to be unrelated to ToM in individuals with ASD or MLD who failed FB tasks, although training children in cognitive flexibility improved their ToM performance at follow-up.



## TABLE OF CONTENTS

<b>Acknowledgements .....</b>	<b>2</b>
<b>Abstract.....</b>	<b>3</b>
<b>Table of Contents .....</b>	<b>4</b>
<b>List of Tables .....</b>	<b>7</b>
<b>List of Figures.....</b>	<b>10</b>
<b>List of Commonly Used Abbreviations.....</b>	<b>11</b>
<b>Chapter 1 .....</b>	<b>12</b>
<b>An Introduction to Autism and ToM.....</b>	<b>12</b>
<i>Psychological theories of autism .....</i>	<i>14</i>
<i>Theory of Mind: a general introduction .....</i>	<i>16</i>
<i>Autism and ToM.....</i>	<i>24</i>
<i>Summary .....</i>	<i>30</i>
<b>Chapter 2 .....</b>	<b>31</b>
<b>The relationship between Theory of Mind and Language.....</b>	<b>31</b>
<i>What is the evidence for a relationship between ToM and language? .....</i>	<i>32</i>
<i>How are language and ToM related? .....</i>	<i>37</i>
<i>Summary .....</i>	<i>45</i>
<b>Chapter 3 .....</b>	<b>47</b>
<b>Study 1: How do they pass and why do they fail? Theory of Mind and</b>	
<b>Language in Children with ASD and Children with MLD.....</b>	<b>47</b>
<i>Introduction .....</i>	<i>47</i>
<i>Method .....</i>	<i>48</i>
<i>Results .....</i>	<i>51</i>
<i>Discussion.....</i>	<i>60</i>
<i>Summary .....</i>	<i>67</i>

<b>Chapter 4 .....</b>	<b>68</b>
<b>The Relationship between Executive Function and Theory of Mind.....</b>	<b>68</b>
<i>EF and ToM in young typically developing children.....</i>	<i>69</i>
<i>EF and ToM in children with developmental disorders. ....</i>	<i>72</i>
<i>Why is there a relationship between EF and ToM?.....</i>	<i>74</i>
<i>Summary .....</i>	<i>84</i>
<b>Chapter 5 .....</b>	<b>87</b>
<b>Study 2: ToM and EF in children with ASD and MLD .....</b>	<b>87</b>
<i>Introduction .....</i>	<i>87</i>
<i>Method.....</i>	<i>89</i>
<i>Results.....</i>	<i>95</i>
<i>Discussion.....</i>	<i>103</i>
<b>Chapter 6 .....</b>	<b>109</b>
<b>ToM and EF Intervention Studies.....</b>	<b>109</b>
<i>Introduction .....</i>	<i>109</i>
<i>Theory of Mind intervention studies. ....</i>	<i>111</i>
<i>Summary .....</i>	<i>122</i>
<i>EF intervention studies .....</i>	<i>124</i>
<i>Summary .....</i>	<i>127</i>
<b>Chapter 7 .....</b>	<b>128</b>
<b>Study 3: A Training Study of Theory of Mind and Executive Function .....</b>	<b>128</b>
<i>Introduction .....</i>	<i>128</i>
<i>Methods.....</i>	<i>129</i>
<i>Results.....</i>	<i>135</i>
<i>Discussion.....</i>	<i>150</i>
<b>Chapter 8 .....</b>	<b>164</b>
<b>Study 4: What Does FB task Success Mean in Individuals with ASD? .....</b>	<b>164</b>
<i>Introduction .....</i>	<i>164</i>
<i>Method .....</i>	<i>172</i>

<i>Results</i> .....	182
<i>Discussion</i> .....	197
<b>Chapter 9</b> .....	<b>204</b>
<b>Study 5: Theory of Mind and Executive Function in Real Life</b> .....	<b>204</b>
<i>Introduction</i> .....	204
<i>Method</i> .....	205
<i>Results</i> .....	208
<i>Discussion</i> .....	214
<b>Chapter 10</b> .....	<b>219</b>
<b>Discussion</b> .....	<b>219</b>
<i>Summary of Results</i> .....	219
<i>Discussion of Results</i> .....	220
<i>General conclusions</i> .....	232
<b>References</b> .....	<b>234</b>
<b>Appendix I Scripts for False Belief Tasks</b> .....	<b>261</b>
<b>Appendix II Questionnaire items</b> .....	<b>262</b>
<b>Appendix III ToM training programme</b> .....	<b>264</b>
<b>Appendix IV EF training programme</b> .....	<b>270</b>
<b>Appendix V Number of children in each trained group passing each ToM task at each time point</b> .....	<b>278</b>
<b>Appendix VI Advanced theory of mind tasks</b> .....	<b>279</b>

## LIST OF TABLES

Table 2.1. Summary of studies on the relationship between social cognition and language .....	32
Table 3.1 Age, gender and language scores for the ASD and MLD groups; mean (s.d.) range.....	49
Table 3.2 Performance on the different FB tasks, split by group. Number (percentage) of participants who passed, failed or failed the control question.....	51
Table 3.3 Number (percentage) participants who passed the Sally-David justification question, having already passed the FB question. ....	51
Table 3.4 Number of inconsistently performing participants with ASD who passed a particular task whilst failing another .....	53
Table 3.5 Number of inconsistently performing participants with MLD who passed a particular task whilst failing another .....	53
Table 3.6 Group characteristics of FB ‘passers’ and ‘failers’ mean (s.d) range.....	55
Table 3.7 Correlations between FB, age and language for the ASD and MLD groups .....	57
Table 4.1. Mean correlations from studies looking at the relationship between ToM and EF tasks in TD children .....	70
Table 4.2. Studies looking at the relationship between ToM and EF tasks in those with developmental disorders. ....	72
Table 5.1. Age and ability of the ASD and MLD groups: mean (s.d.).....	90
Table 5.2 ToM and control task performance split by group (percentage passing or mean; s.d).....	96
Table 5.3 EF task performance split by group.....	97
Table 5.4 Pearson’s correlations between EF tasks. ASD group (N = 27) above diagonal in bold, MLD group (N=21) below diagonal.....	99
Table 5.5. Correlations between age, ability, ToM and EF for the ASD group (N = 27) .....	100
Table 5.6. Correlations between age, ability, ToM and EF for the MLD group (N = 21) .....	100

Table 5.7 Pearson's correlations between the ToM and EF measure for the ASD (N = 27) and MLD (N =21) groups (ASD above diagonal in bold).....	101
Table 6.1. ToM training studies.....	116
Table 7.1 CA and ability of the six experimental group: mean (s.d.).....	130
Table 7.2 Overview of project design.....	131
Table 7.3 Tasks used at each timepoint .....	131
Table 7.4. Baseline scores on the ToM total score, the Eyes task and EF measures for the six experimental groups. Mean (s.d.).....	136
Table 7.5. Mean percentage of ToM tasks (s.d) passed at each stage for each experimental group. ....	137
Table 7.6 Percentage of children in each group who improved on the proportion of ToM tasks passed between pre-test and post-test or pre-test and follow-up (percentage improving on a task other than the unexpected transfer task).....	139
Table 7.7 Performance on the Eyes task at pre-test and follow-up: mean (s.d.) .....	140
Table 7.8 Scores on Card Sort variables for the six experimental groups at pre-test, post-test and follow-up .....	141
Table 7.9 Performance on Go-NoGo and Trails at pre-test, post-test or follow-up .	144
Table 7.10 Percentage of children improving or not between pre-test and follow-up on the Trails task.....	144
Table 7.11 Spearman's correlations between the age and ability measures and the measures of success during the EF training. ASD group in bold. ....	147
Table 7.12. Mean (s.d) scores on the 3 questionnaire subscales for each experimental group .....	149
Table 8.1. Age and ability measures for the two groups; mean (s.d.) .....	173
Table 8.2. Age and ability of ability matched subgroups; mean (standard deviation) .....	174
Table 8.3 A summary of the tasks used in Study 4, and the reason for their inclusion .....	175
Table 8.4 Performance on the Eyes task and the Eyes vocabulary test in the ASD and MLD groups; mean (s.d.).....	183
Table 8.5 Percentage of children in each group able to do each aspect of the Think task .....	185
Table 8.6 Percentage of children in each group able to do each aspect of the Not think task .....	185

Table 8.7 Mean scores on the Cartoons and Stories for the ASD and MLD groups; mean (s.d).....	186
Table 8.8 Percentage of each group who passed each FB task, passed the justification question or failed the control question.....	187
Table 8.9 Mean reading times for each story condition as a proportion of total reading time: mean (s.d) .....	189
Table 8.10 Scores for the Picture Sequences task for the ASD and MLD groups: mean (s.d) .....	192
Table 8.11 Correlations between task aggregates and ability measures in the ASD and MLD groups (ASD in bold).....	195
Table 8.12 Correlations between groups of ToM tasks and the non-ToM control tasks. ASD (N=19) above the diagonal and in bold, MLD (N=22) below diagonal.....	196
Table 9.1 Characteristics of participants in the ASD and MLD groups: mean (s.d.)	206
Table 9.2 Mean score on the 3 questionnaire scales. Mean, (s.d.), range.....	208
Table 9.3 Scores for FB passers and failers on age, verbal ability and non-verbal ability, and the ToM, EF and Sociability questionnaire scales (mean (s.d.) for ability and age, mean, (s.d.) range for scales).....	209
Table 9.4 Number of participants in each group who scored at or above the median for the questionnaire subscales .....	210
Table 9.5. Mean scores for the card sort variables for the ASD and MLD groups; mean (s.d.).....	211
Table 9.6 Correlations between questionnaire subscales and card sort variables in the ASD and MLD groups (values with Raven's partialled out). ASD in bold. ....	212
Table 9.7 Correlations between the questionnaire subscales and ability measures (ASD group in bold). ....	214

## LIST OF FIGURES

Figure 2.1 Tager-Flusberg's model of ToM and Language development (adapted from Tager-Flusberg, 2002) .....	45
Figure 3.1. Percentage of each group who passed 0, 1, 2 or 3 FB questions.....	53
Figure 3.2. Scatter plots of FB passers and failers on TROG VMA v CA.....	58
Figure 3.3 Scatter plots of passers and failers on BPVS VMA v. CA.....	59
Figure 7.1: Graphs showing proportion of ToM tests passed at pre-test, post-test and follow-up in the experimental groups.....	138
Figure 7.2 Percentage of each group improving on the Eyes task between pre-test and follow-up.....	140
Figure 7.3. Performance of each experimental group on the Card Sort aggregate at each time point.....	142
Figure 7.4 Percentage of each group improving on the Card Sort aggregate between pre-test, post-test and follow-up .....	143
Figure 7.5 Percentage in each group who improved on the Go-NoGo task between pre-test and post-test .....	145
Figure 8.1 Illustration of the interaction between cartoon type and group for VMA-matched subgroups .....	187
Figure 8.2 Performance on the Unexpected Stories task .....	191
Figure 8.3. Percentage passing both sequences correctly in each condition, split by group .....	193
Figure 8.4. No. of participants who performed better on each condition in the ASD and MLD group. ....	194

<b>LIST OF COMMONLY USED ABBREVIATIONS</b>
--

ADHD = Attention Deficit Hyperactivity Disorder

ASD = Autistic Spectrum Disorder

AS = Asperger's Syndrome

BPVS = British Picture Vocabulary Scale

CA = Chronological Age

CPM = Raven's Coloured Progressive Matrices

DCCS = Dimensional Change Card Sort task

DS = Down Syndrome

EF = Executive Function

FB = False Belief

MLD = Moderate Learning Difficulties (i.e. non-autistic)

TD = Typically Developing

ToM = Theory of Mind

TROG = Test for Reception of Grammar

VMA = Verbal Mental Age

WCST = Wisconsin Card Sort Task



## **CHAPTER 1**

### **AN INTRODUCTION TO AUTISM AND TOM**

The first chapter of this thesis provides an introduction to autistic spectrum disorders, a brief description of three major cognitive theories of autism and an overview of current research into theory of mind, both in normal development and in autism.

In 1943, Kanner described a group of children who had a puzzling combination of symptoms – an obsessive desire for sameness, echolalia, lack of social responsiveness, oversensitivity to stimuli and restricted interests, combined with a good memory and a seemingly good cognitive potential (Kanner, 1943). His was the first description of autism as a disorder, and his description is strikingly similar to what is understood by ‘autism’ today. Autism, according to both ICD-10 (WHO, 1993) and DSM-IV (APA, 1994) diagnostic criteria, is characterised by early (evident before the age of 3) and persistent impairment in social interaction, communication and imagination, with restricted and repetitive interests and activities. The deficits in socialisation, communication and imagination are known as the ‘triad of impairments’ (Wing & Gould, 1979). Wing and Gould also introduced the concept of an autistic spectrum, on which autism falls, along with other pervasive developmental disorders such as Asperger’s Syndrome and pervasive developmental disorder-not otherwise specified (PDD-NOS). Throughout this thesis, the term autistic spectrum disorder (ASD) will be used to refer to groups of children who have a diagnosis of either autism, Asperger’s Syndrome or an autistic spectrum disorder.

The autistic spectrum includes all levels of intellectual ability, ranging from severely impaired individuals who never learn to speak, through to very high-functioning people capable of obtaining university degrees and more. Pervasive developmental disorders affect approximately 60 per 10,000 children (Baird et al., 2000; Chakrabarti & Fombonne, 2001). Precise estimates differ according to the recency of the study (with more recent studies finding a higher prevalence), the age group studied, and whether the study relies on diagnostic records or uses more active ascertainment. Recent estimates are substantially higher than had been previously

thought and appear to be on the increase (Powell et al., 2000), but whether this is due solely to increased awareness and changes in diagnostic practice or whether it is also due to a genuine increase in prevalence is not known.

Although autism is defined behaviourally, it is a biological disorder and is considered to be one of the most heritable developmental disorders, with estimates of heritability exceeding 90% (Bailey et al., 1995). Levels of concordance between monozygotic twins range between 60 and 80%, and the risk of a sibling of a child with autism themselves having an ASD is between 3 and 6% (Bailey et al., 1995; Rutter, Silberg, O'Connor, & Simonoff, 1999). The pattern of heritability makes it clear that autism is not a single gene disorder, but instead is likely to be caused by a small number of interacting genes, perhaps combined with some environmental insult (since the concordance rate amongst monozygotic twins is not 100%; (Gillberg & Coleman, 2001). It is not clear which environmental events (e.g. drugs or viruses) could be significant and there is little evidence to date.

Autism is far more prevalent in males than in females, and this discrepancy becomes more pronounced at the high-ability end of the spectrum (Wing, 1981). The reason for this male bias is not known. It is possible that the current diagnostic criteria may identify ASDs more successfully in men than in women, or that more severe brain involvement is necessary for a female to appear to have an ASD.

The biological basis of autism is now widely accepted (Gillberg & Coleman, 2001). This is, however, a relatively recent development. Bettelheim (1956) suggested that children became autistic as a result of an unloving environment – the ‘refrigerator mother’ theory. This idea was highly influential at the time in forming professionals’ attitudes and the treatment of autism. It is now clear that children can be badly neglected and mistreated and yet not become autistic (Clarke & Clarke, 1976). However, studies of Romanian adoptees (who were severely deprived in the early years of life) have shown a subgroup of those who were kept in institutions for 2-3 years do display some autistic traits in early and middle childhood, including less pretend play and fewer references to mental states than a control group (Kreppner, O'Connor, Dunn, & Andersen-Wood, 1999). However, these cases of severe institutionalised sensory and emotional deprivation are quite distinct from the normal range of parenting experienced by children, and appear to result in a form of quasi-autism with a different course, in which the symptoms are somewhat attenuated as the child grows up.

## **Psychological theories of autism**

Whilst autism is clearly a biological disorder and is described on a behavioural level, cognition links these two levels of explanation and is crucial both in understanding the behaviour of individuals with ASD and in pinpointing potential areas of the brain in which abnormalities may be found, thereby providing pointers as to which genes might be involved. There are several current cognitive theories of autism that will be outlined in brief. This is not an exhaustive review of all theories of autism, but rather a selective discussion of those that may be relevant to the experimental questions and research reported in this thesis. The theories discussed broadly fall into two types, social theories (e.g. theory of mind) and non-social theories (e.g. executive function and central coherence). There are other theories of autism, such as Hobson (2002), who suggests that the core deficit in autism is in interpersonal relationships, which in turn leads to difficulties such as theory of mind deficits. However, these are not relevant to the studies in this thesis and will not be discussed. Theory of mind (ToM) will be the main focus of this thesis, and as such will be discussed last and in the greatest detail.

### **Executive Function**

Executive Function (EF) is an umbrella term for higher order cognitive function, covering inhibitory control, set shifting and flexibility, planning and working memory. The inflexible and repetitive behaviour seen in many children with ASD, and the similarity in some cases to the behaviour seen in patients with frontal lobe dysfunction, led researchers to investigate their executive functioning. A wide range of studies (e.g. Hughes & Russell, 1993; Hughes, Russell, & Robbins, 1994; McEvoy, Rogers, & Pennington, 1993; Ozonoff, Pennington, & Rogers, 1991) have found that children with autism show deficits in tests of EF when compared to typically developing (TD) controls, particularly on tasks of set shifting and flexibility (Ozonoff, 1997), but also on tasks of planning and fluency (Sergeant, Geurts, & Oosterlaan, 2002). Some researchers have proposed that executive function deficits may be the primary deficit in ASD, and thus account for difficulties seen in other areas (Ozonoff et al., 1991; Russell, 1997b). However, it is not clear whether such difficulties can discriminate between children with ASD and other groups with developmental disorders such as Attention Deficit Hyperactivity Disorder (ADHD), and therefore whether the deficits are specific to autism (Sergeant et al., 2002). The

majority of studies looking at EF deficits in clinical groups such as those with ASD or ADHD compare a single clinical group to controls rather than comparing between clinical groups. There is some evidence from a single study that children with autism perform worse on a test of flexibility and set shifting (the Wisconsin Card Sorting Test, WCST; (Heaton, 1981) than do children with ADHD (Ozonoff & Jensen, 1999), whilst both clinical groups perform worse than controls on this task across a range of studies (Sergeant et al., 2002). However, there is also evidence that young children with ASD do not have EF problems, which make claims that EF is the primary deficit in autism problematic (Griffith, Pennington, Wehner, & Rogers, 1999). This area needs further investigation, to see if there is a specific profile of EF weaknesses in ASD, as opposed to other clinical disorders, and to see whether it is plausible that this could be the primary deficit. Executive function, and its relationship to theory of mind, will be discussed further in chapter 4.

### **Weak Central Coherence**

Anecdotal accounts of children with autism often refer to their ability to notice tiny changes in the environment, and to repeat verbatim long stretches of dialogue from videos; whilst appearing not to understand either what they are saying or the underlying gist of the story. Frith (1989) proposed that these features of autism are part of a tendency towards detail focused processing. Typically developing children and adults have a strong tendency to extract a global meaning from information – so-called ‘central coherence’ - for example remembering the gist of a story whilst forgetting the details. Frith (1989) suggests that this fundamental feature of information processing is disturbed in autism, and that this could account for some of the ‘islets of ability’ seen in autism, particularly in those with savant skills. Shah and Frith found that individuals with autism showed highly superior performance on the Embedded Figures Test, (which requires the participant to pick out hidden figures from a picture) and also showed superior performance on the Block Design task (Shah & Frith, 1983; Shah & Frith, 1993), both tasks in which local, rather than global, processing might be advantageous. Weak central coherence is proposed to be a cognitive style, more prevalent in people with ASD, just as a tendency towards strong central coherence is more prevalent in normal individuals. The tendency towards this cognitive style predicts a pattern of cognitive strengths and weaknesses, as opposed to the ToM and EF theories, which suggest only weaknesses. However, it is not clear

what mechanism might underlie this local bias. Alternative explanations for the characteristic pattern of strengths and weaknesses seen in autism have included enhanced perceptual functioning (Mottron & Burack, 2001) or a reduced ability to generalise (Plaisted, 2001).

### **Theory of Mind**

Between the ages of 3 and 4 years, TD children become able to pass simple tests of ToM – assessing their ability to attribute thoughts and beliefs in order to predict and explain behaviour. Baron-Cohen, Leslie and Frith (1985) proposed that the deficits in socialisation, communication and pretend play seen in autism could be the results of a failure to develop this understanding of mental states or ‘mindreading’. This theory is the main focus of this thesis, and as such research in this area will be discussed in detail below. First, the study of ToM in autism will be placed in context through a discussion of ToM research in normal development. Following that, ToM research will be discussed specifically with reference to ASD, and the implications of the postulated ToM deficit in ASD will be discussed. Throughout this text the terms ‘ToM’, ‘mentalising’ and ‘understanding of mind’ will be used to refer to the ability to attribute independent mental states such as beliefs.

### **Theory of Mind: a general introduction**

The concept of a ‘theory of mind’ was first introduced by Premack (Premack & Woodruff, 1978) with reference to chimpanzees. By a ToM, Premack meant the ability to represent mental states – and in the context of chimpanzees this was postulated to be an innate cognitive mechanism, rather than any sort of conscious ‘theory’. Work with young TD children followed, using a test which has now become standard and which will be referred to throughout this text as the ‘unexpected transfer’ false belief (FB) task. In one well-known variation of this task (Baron-Cohen et al., 1985), a protagonist (Sally) puts a marble into a basket and leaves the room. Another character (Anne) moves the marble from the basket into a box whilst Sally is absent. Sally then returns, and the child is asked where Sally will look for her marble. Wimmer and Perner (1983) found that whilst most TD 3-year-olds failed simple tasks of this type, almost all 4-year-olds could pass, indicating that they could distinguish between the protagonist’s (mistaken) belief and the reality. Another frequently used FB task (Hogrefe, Wimmer, & Perner, 1986; Perner, Leekam, &

Wimmer, 1987) will be referred to as the ‘deceptive box’ task throughout this text. In this task the child is shown a familiar box (for example a Smarties box) and is asked what the box contains. Typically children reply ‘Smarties’. The box is then opened and they are shown that it actually contains a pencil. The box is closed again and they are asked ‘If your friend came in, what would he think was in the box?’ (the *other* FB question). They are then also sometimes asked ‘Before I opened the box, what did you think was inside?’ (the *self* FB question). Both these tasks test the representation of mental states by requiring the child to attribute a belief that is different to their present belief, and indeed to reality.

There has been much debate over what it means for a young child to either pass or fail a FB task (e.g. Bloom & German, 2000), with many researchers arguing that casting 3-year-olds as lacking ToM ignores the social sophistication that they demonstrate through speech (where they can contrast mental states well before their fourth birthday; Bartsch and Wellman, 1995), pretend play (Leslie, 1987), and prosocial behaviour (where 2-year-olds are known to show empathy; Zahn-Waxler, Robinson and Emde, 1992). Some authors have argued that FB tasks place demands (for example executive, or linguistic) on the child that have nothing to do with having or lacking a ToM, but which account for the failure of most 3-year-olds (Hughes & Russell, 1993; Siegal & Beattie, 1991). Task manipulations such as asking the child where the protagonist will look *first* have enabled children younger than 4 years to pass these tests (Siegal & Beattie, 1991). However, a recent meta-analysis of FB tasks (Wellman, Cross, & Watson, 2001) found that even with these task manipulations, children younger than 3 overwhelmingly still fail, and that the improvement seen is generally due to their performance moving from being at below chance to chance level – not very strong evidence of the ability to attribute independent mental states. Wellman et al. (2001) found that the improvement in performance on FB tasks between the ages of 3 and 4 is consistently seen across a wide range of studies. Whilst it is indisputable that 3-year-olds and even 2-year-olds have some awareness of other people’s mental states, it also seems that in most cases they genuinely do fail to attribute false belief in standard tasks. Explaining this apparent discrepancy between real life behaviour and experimental performance has been a major challenge for researchers investigating ToM in TD children.

Over the past 20 years the development of ToM has received a great deal of attention from researchers (see Flavell, 1999 for a review), so much so that it is

beyond the scope of the present chapter to conduct an exhaustive summary. I will briefly outline the main theories of how a TD child might acquire a ToM, and will then go on to discuss ToM research in autism.

Theories of how a child acquires a ToM fall into two broad types, domain-specific and domain-general. The domain-specific theories postulate that specific advances in a child's social understanding allow them to develop a ToM (as distinct from general cognitive advances), whilst the domain-general theories suggest that there is nothing special about a ToM, but that general cognitive advances allow a child to pass FB tasks and presumably to gain an understanding of minds. The main domain-specific theories are simulation theory, modularity theory and the 'child as scientist' theory. This last theory could be put into either the domain-specific or domain-general group, since it postulates that the child uses domain-general mechanisms (those of theory formation) to develop a domain-specific theory of mind.

The difference between passing FB tasks and actually having a 'theory of mind' is an important one to keep in mind - other theories, not reviewed here (but discussed later in this thesis), propose explanations for why a child might fail FB tasks, but do not address the development of the underlying understanding of mind.

### **Domain-specific theories**

Simulation theory (e.g. Harris, 1992) suggests that children can use their own minds as models for understanding others, and in particular for making predictions about what others will think and do. We know what we would do in a particular set of circumstances, and we are able to project that onto others, as if we were re-running our own system offline. Some argue this involves an introspective access to our own mental states (Harris, 1992), whilst others argue that it is a more automatic non-conscious process (Gordon, 1986). Harris suggests that the simulation processes are set in motion by inbuilt mechanisms for establishing joint attention and a joint emotional stance. The ability to model other's behaviour has been proposed to develop in young children through role-taking (for example through pretend play). Harris (1992) argues that some forms of simulation are more complex than others, and in particular that running a simulation which requires imagining beliefs that run counter to the child's own (i.e. a false belief) is complex and requires imaginative flexibility, lack of which prevents younger children from passing FB tasks. Harris's account in particular would predict that children should be able to report on their own

mental states before they can reflect on others'. There is little evidence for this. Wimmer, Hogrefe and Perner (1988) found that there was no difference in the age at which a child became able to predict another's FB and the age at which they could recall their own. Harris counters this by claiming that recalling past beliefs which the child now knows to be false involves the same imaginative construction of a counterfactual state as predicting another's belief. However, work by Flavell and colleagues (Flavell, Green, & Flavell, 1995, 2000; Flavell, Green, Flavell, & Grossman, 1997) appears to show that access to and certain aspects of an awareness of one's own mental states (for example, the ability to report on one's own stream of consciousness) arrives surprisingly late (certainly after the age of 5), even when there is no conflict between the belief state and reality.

Modularity theorists suggest that a 'theory of mind mechanism', perhaps a module, or series of connected modules, have an innate basis in the brain, and that neurological maturation accounts for the transition seen between the ages of 3 and 4. Leslie (Leslie & Roth, 1994; Leslie, 1994) suggests three domain-specific mechanisms, which mature in succession. The first, called Theory of Body (ToBy) develops very early, in the first year. It allows infants to recognise that agents are able to move on their own. This is as opposed to non-agents, which must be acted upon. The next mechanism, called ToM Mechanism *system 1* (ToMM<sub>1</sub>) deals with intentionality of agents, i.e. recognising that agents act upon intentions such as desires. This, Leslie suggests, matures later in the first year. The final mechanism, ToMM *system 2* (ToMM<sub>2</sub>) develops during the second year of life and allows the child to represent propositional attitudes such as pretending that, believing that or thinking that, termed M-representations by Leslie. This is obviously much earlier than the age at which TD children pass tests of FB. Leslie then postulates that difficulties with another, non-ToM mechanism, the selection processor (which he defines as a general executive function required in many situations to inhibit salient but unwanted responses) cause 3-year-olds to fail FB tasks, rather than difficulties in mentalising. Maturation of the selection processor allows 4-year-olds to demonstrate their M-representational ability.

Baron-Cohen (1995) also suggests a set of innate modular mechanisms. He suggests that the earliest building blocks for ToM that the infant possesses are the Intentionality Detector (ID), which allows the child to interpret actions in a goal-oriented way, and the Eye-Direction Detector (EDD) which detects the presence of



eyes, computes what the eyes are directed towards, and then infers that if another organism's eyes are directed at something then that organism sees that thing. Both ID and EDD allow dyadic representations only – they represent a relation between two objects, the agent and an object, or the agent and self. *In order to share an experience* with another person, a triadic representation is necessary, and for this another module – the Shared-Attention Mechanism (SAM) – is necessary. This uses information about another agent's perceptual state – for example information detected through EDD, to compare that perceptual state with the child's own perceptual state, and thus to share attention. The final mechanism, the ToM Mechanism (ToMM) is necessary for the child to represent epistemic mental states (as opposed to the volitional or perceptual mental states which the other three mechanisms allow) such as pretending, thinking, believing and knowing. It also ties together all the different mental-state concepts into one coherent understanding.

These modular explanations do not completely discount the role of experience, however, they postulate that the mechanisms for a ToM are innate, and whilst experience may be necessary to trigger the operation of the mechanisms, it will not determine their nature. This contrasts with the 'child as scientist' view (e.g. Gopnik & Wellman, 1994) which argues that children's knowledge about the mind goes through a number of theoretical stages, each of which is superseded by a new theory when evidence and experience lead the child to reject the earlier theory as inadequate. As stated above, this theory could be classed as either domain-specific or domain-general, since the processes which enable a child to acquire a ToM are postulated to be domain-general, whilst the final product (the 'theory of mind') is domain-specific. The child is postulated to have a succession of theories, each of which is replaced by a new theory as counter-evidence is collected. In elucidating how a child might come to develop ToM, Bartsch and Wellman (1995) describe a three-step developmental sequence. The first stage (at age 2) is a 'desire psychology', which they suggest involves a mentalistic but non-representational view of the mind. The child understands that people have inner experiences connected to objects (e.g. wanting them or seeing them) but does not understand that people mentally represent objects as being a certain way. The next stage, at around age 3, is heralded by the child beginning to talk about beliefs and thoughts as well as desires, and involves an understanding that mental representations exist which can be false or true, and can differ between people. However, despite this understanding children continue to

explain actions by desires rather than beliefs. Bartsch and Wellman term this a 'desire-belief psychology'. At around the age of 4, children begin to understand that people's thoughts and beliefs, as well as their desires, influence their actions, and thus they acquire a 'belief-desire psychology'. Experience is therefore crucially important in children's ToM development in providing young children with information that cannot be accounted for by their present theory, and that will therefore eventually lead to conceptual change. In its strongest form, the 'child as scientist' theory does not postulate any innate mechanism, and therefore presumably children with autism would be assumed to have general difficulties in theory-formation, or lack the necessary experience to form their theory, as the explanation for their difficulties in ToM. However, the question must be asked, if there is no innate basis for the understanding of mind, then why do all TD children come up with the same theory and within the same time frame, despite great variation in experience, ability and culture (Avis & Harris, 1991)? If we assume that there is an underlying innate basis on which the theory is formed (as in fact recent versions of the 'child as scientist' theory do (Gopnik, Capps, & Meltzoff, 2000)), then the account begins to blur more with a modular approach, and the theory loses some of its distinctiveness. Gopnik, Capps and Meltzoff (2000) argue that their account differs from modular accounts because the innate 'theory' they suggest is hard-wired in the brain is a starting point, rather than the final architecture of theory of mind, and because they interpret Fodor (1983) as stating that information from other cognitive systems cannot be used in the construction of modular representations. However, as Coltheart (1999) points out, many of the modular theories of ToM do not take a strong Fodorian approach and allow for the role of experience in forming ToM development – indeed, given the evidence of delay from those with perceptual deficits (e.g. deaf children of hearing parents; Peterson and Siegal, 1995), it is hard to maintain an alternative stance.

It seems, then, that the latest forms of all three of these theories postulate an innate, specialised system that underpins the development of a ToM. The theories vary on when this innate system becomes important, and how sophisticated they suggest it is, and how much interaction between this system and the environment is necessary in order to form a fully functioning ToM.

### **Domain-general theories**

More domain-general theories argue that young children's failure on FB tasks is the result of limitations in general information processing. So for example, the Cognitive Complicity and Control (CCC) theory (Frye, Zelazo, & Burack, 1998; Frye, Zelazo, & Palfai, 1995; Zelazo, Burack, Benedetto, & Frye, 1996; Zelazo, Burack, Boseovski, Jacques, & Frye, 2001) argues that passing a FB task reflects the 4-year-old's new ability to use embedded rule based reasoning, meaning reasoning with a setting condition, an antecedent and then a consequence. They base this argument on the observed relationship between the Dimensional Change Card Sort task (DCCS; in which the child is required to sort cards using an *if-if-then* rule structure) and performance on FB tasks. For example, in the DCCS a child is required to sort coloured cards according to either shape (the *Shape* game) or colour (the *Colour* game). The reasoning structure, according to Frye et al., therefore goes like this: *If* we are playing the Shape game and *If* the card is a triangle, *Then* place in left hand box. Frye et al (1995) suggest that a FB task can be described in the same form as follows: *If* Sally *If* looking for marble *Then* will look here. However, Perner (Perner, Stummer, & Lang, 1999) suggests that the FB scenario could just as easily be cast as: *If* Sally is looking for her marble *Then* will look here – only an *if-then* reasoning structure which 3-year-olds can easily master. In addition, whether this logical reasoning structure also applies in everyday understanding of minds is less evident.

De Villiers (De Villiers, 2000) argues that a ToM comes about as a result of linguistic gains, specifically complement syntax (e.g. 'he thinks that the horse is in the garden', where 'the horse is in the garden' may be true or false without affecting the truth value of the whole sentence). She uses evidence from deaf children of hearing parents – who typically do not pass FB tasks until much later than TD children and who have impoverished access to language in their early years - to support this. This account is in principle domain-general, since linguistic gains rather than specific gains in ToM lead to a child passing FB tasks. However, in practice complement syntax only applies to mental state and communication verbs in English, and so the claim is not testable in other domains. It is also not clear whether this account is one which explains competence in understanding other minds in every day life, or just in ToM tasks. To my knowledge, the evidence for ToM competence (or not) in the every day life of deaf FB failers has not been examined. It is also possible that deaf children have delayed ToM due to their relatively impoverished social environment, in

particular the lack of access to conversation (Peterson & Siegal, 1995; Woolfe, Want, & Siegal, 2002) – deaf children of hearing parents are likely to miss out on a huge amount of communication in their early years.

Perner (Perner & Lang, 1999, 2000) argues that ToM is in fact an integral part of executive functioning, and that the child comes to understand the representational nature of the mind through learning to inhibit executive actions, which he argues requires the child to understand the existence of causal schemata. Understanding the existence of these schemata as entities that make one act in a certain way is to understand them as representations with causal power, and Perner argues that this understanding of representation is what makes it possible for a child to represent mental states such as thoughts and beliefs. Therefore the development of ToM is not a domain-specific achievement, but in fact part of a wider cognitive leap in understanding. This, like Zelazo's (1996; 2001) theory, accounts for children with autism's observed difficulties in both EF and ToM, and the relationship between the two seen in young TD children (see chapter 4).

Others (as I have discussed above) have argued that the executive demands of FB tasks cause failure (Hughes & Russell, 1993), or that the way the question is asked causes young children to misunderstand (Siegal & Beattie, 1991). These claims will not be discussed in any detail here (although they will be discussed in chapters 2 and 4), as they do not make predictions about the development of a genuine understanding of mind, only about passing FB tasks, and therefore still leave open the question of how an understanding of mind, rather than the ability to pass FB tasks, develops. In addition, research by Hughes (1998a) and the Wellman et al. (2001) meta-analysis suggest that it is unlikely that these task-specific factors can account for the transition in performance on ToM seen between the ages of three and four.

The relationship between a ToM and other cognitive domains is an area of growing interest. Some of the theories as to the relationship between ToM, language and EF have been touched upon above. These, along with other alternatives, will be discussed in detail in chapters 2 and 4, and investigated in studies 1, 2 and 3. Understanding the relationship between ToM, EF and language may give us important insights into the way that the understanding of mind develops. This is particularly relevant in children with ASD, who have well documented difficulties in ToM.

## **Autism and ToM**

Investigations into the understanding of mind in children with autism date back to Baron-Cohen, Leslie and Frith's (1985) seminal paper, which found that 80% of children with autism failed simple 'unexpected transfer' FB tasks, as compared to fewer than 20% of verbal age matched controls (who had Down Syndrome). In the seventeen years since then a large number of studies, using a wide range of tasks, have assessed the understanding of minds in children with autism (see Baron-Cohen, 2001 for a review). These tasks have included tests of deception, tests of recognising mental state terms and of recognising the functions of the brain, as well as more complex tests such as 2<sup>nd</sup> order FB tests (in which the participant is asked 'what does John think that Mary thinks?'), stories about complex mental states, and identifying mental states from the expression in the eyes. Although there have been exceptions (Bowler, 1992; Charman & Lynggaard, 1998; see Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998 for a meta-analysis), the great majority of studies have found a deficit in the understanding of mind in children with autism when compared to verbal-age matched controls (Baron-Cohen, 2001). The evidence to date suggests that children with autism specifically lack the ability to meta-represent and so are unable to attribute mental states or propositional attitudes, whilst being able to understand non-representational states such as wanting and seeing.

How can the various theories of ToM deal with this proposed deficit in autism? It is perhaps most straightforward for modularity theorists such as Leslie or Baron-Cohen, who argue that children with autism have a specific deficit in the ToM mechanism – in Leslie's account in ToMM<sub>2</sub>, and in Baron-Cohen's account in the Shared Attention Mechanism (SAM) and as a secondary effect in ToMM. Simulation theorists (e.g. Harris, 1992) argue that children with autism may lack the innate mechanism that enables joint attention to be established, and from which simulation processes are postulated to start. However, as Harris himself points out, this cannot explain the selective difficulties that children with autism have with representational mental states, as opposed to volitional or perceptual mental states such as seeing and wanting. A strong 'child as scientist' account, which postulates no innate basis for a ToM would have to predict general difficulties in theory forming in autism, one result of which is the lack of a ToM. However, in fact recent accounts which address the ToM deficit in autism from a theory building perspective postulate an innate starting

state 'theory' which children are born with and which allows them to make predictions, interpret and maybe even explain what they see around them (Gopnik et al., 2000). The child then builds on this innate theory to develop a fully functioning representational ToM by the age of 4, and the child with autism either lacks this innate theory, or has a radically different theory. However, as discussed briefly above, the innate 'theory' in account does not appear to differ so much from a neo-Fodorian module as outlined by Coltheart (1999; in which the fundamental property of a module is simply domain specificity). After all, the development of a module for reading (for example) requires substantial external input before it is functional, but this in itself is not a reason for declaring something non-modular. There is no apparent reason why the innate 'theory' that Gopnik et al (2000) suggest underlies ToM development should not be called a module in the sense meant by Coltheart (1999), and once we accept the existence of an innate module underlying theory development this view and that of the modularity theorists draw closer together. Since simulation theorists also postulate an innate basis for ToM development (Harris, 1992), there is some level of agreement amongst all these theories that ToM has an innate basis and it is this innate basis which is impaired in ASD.

Other domain-general theories predict that children with ASD's difficulties in ToM will inevitably be associated with difficulties which would show up in other tests – in the case of the CCC, the problems would be in rule-based reasoning (Zelazo et al., 2001), whilst in De Villiers' (2000) account language difficulties would be inevitably associated with deficits in ToM. Perner's (Perner et al., 1999) account would predict general difficulties in meta-representation, not just in the representation of mental states. The implication of all of these is that the deficit seen in autism is not in fact specific to ToM, but is indicative of more pervasive cognitive difficulties. There is some evidence that this is not the case – for example, Leekam and Perner (1991) found that children with autism did not have deficits on a test of physical, rather than mental, representations.

The ToM theory of autism has been highly influential over the last 17 years. On a descriptive level, it can account for many of the difficulties seen in ASD. We make sense of the social world we inhabit in terms of mental states, and a lack of an ability to conceptualise those mental states will lead to profound difficulties in socialisation and communication. The possible link between a ToM deficit and these difficulties is perhaps easiest to conceptualise in the case of communication deficits,

which are evident even in those individuals with ASD who appear to have very good language skills. Individuals with ASD are prone to taking people literally – for example, when asked on the phone ‘Can you get your mum?’ a child with autism may well answer ‘Yes’ and put the phone down! A TD child will understand very early on that the question in fact is a request to get their mother, rather than a question about their capability to do so. However, in order for the child to understand this it is necessary for them to understand the intention that is the mental state behind the question, rather than taking the words at face value. There is plenty of experimental as well as anecdotal evidence that individuals with ASD have difficulty understanding non-literal or figurative speech (Happé, 1993, 1994; Minshew, Goldstein, & Siegal, 1995). It is not hard to see how such profound difficulties in communication will lead to difficulties in social interactions: without recourse to mental states in order to explain actions and motivations, other people’s behaviour will seem chaotic and senseless to the individual with ASD. In addition, a lack of understanding of mental states may directly cause problems in socialisation, since children who have no understanding of other’s minds will not seek to share attention or enjoyment, and will have poor understanding of emotions.

The ToM theory of autism has enabled researchers and professionals to predict very precisely the behaviours that children with autism will have difficulties with (e.g. understanding complex emotions, attributing representational mental states) and those on which they will be unimpaired, such as understanding simple emotion and attributing non-representational mental states (Baron-Cohen, Spitz, & Cross, 1993). It has also provided a starting point for examining possible brain abnormalities by functional neuroimaging, and for identifying areas of the brain that may be compromised in autism (Fletcher et al., 1995; Gallagher et al., 2000; Happé et al., 1996). This, in conjunction with bottom-up molecular genetic investigations, may in the long term lead to the identification of specific genes that are important in the aetiology of autism.

On a more practical level, the theory has led to the development of tools for early screening such as the Checklist for Autism in Toddlers (CHAT: Baird et al., 2000; Baron-Cohen, Allen, & Gillberg, 1992; Charman, Baron-Cohen, Baird et al., 2001), which has had some success in identifying 18 month old children who will go on to be diagnosed with autism. It has also led to the development of intervention programmes that target difficulties in ToM (Hadwin, Baron-Cohen, Howlin, & Hill,

1996, 1997; MacGregor, Whiten, & Blackburn, 1998; Swettenham, 1996; see Swettenham, 2000 for a review; Swettenham, Baron-Cohen, Gomez, & Walsh, 1996), often by identifying those areas that children with autism have difficulties with (e.g. understanding mental representations) and exploiting areas in which they are unimpaired (e.g. understanding physical representations). No other theory of autism has had so much influence in both research and clinical practice. Despite this, there are aspects of autism which a ToM deficit cannot account for, in particular repetitive behaviour, and the uneven cognitive profile and 'islets of ability' often seen in people with autism.

### **Specificity and universality**

Perhaps surprisingly for a theory that has had so much influence on the field, it is not clear whether the ToM deficit is specific to autism, nor whether it is universal to all children with autism. Certainly individuals with autism are not the only group to demonstrate difficulties with ToM tasks. There is evidence for ToM task failure in deaf children of hearing parents (De Villiers, 2000; Peterson & Siegal, 1995), blind children (Minter, Hobson, & Bishop, 1998; Peterson, Peterson, & Webb, 2000), people with schizophrenia (Corcoran, 2000) and stroke patients (Happé, Brownell, & Winner, 1999). However, the real life social and communication deficits seen in these groups are different from those in autism in that they tend to be less severe (often resolving with time) and either acquired (in the case of schizophrenia and stroke patients) or due to impoverished perceptual input (in the case of deaf and blind children).

There have also been some studies that have found deficits in performance on ToM tasks in children or adults with non-autistic moderate learning difficulties (Ashcroft, Jervis, & Roberts, 1999; Zelazo et al., 1996). The case of children with moderate learning difficulties (MLD) may be more relevant to the issue of specificity of the ToM deficit in autism, since it is very unlikely that these children have an acquired deficit in ToM, and they have no perceptual problems that would lead us to predict a ToM delay. They may be generally intellectually delayed, but we would still expect them to pass FB tasks at the same developmental level as do other, TD children. Children with MLD typically do not demonstrate the profound social and communication deficits seen in ASD, and therefore any evidence for ToM deficits in this group would call into question the role of ToM as a causal factor underlying key



diagnostic features seen in autism. Yirmiya et al. (1998), in a meta-analysis of 40 studies of ToM performance (using standard FB tasks) in individuals with autism, MLD and TD children, found that individuals with autism differed significantly (in that more of them failed ToM tasks) from both those with MLD and TD young children. They also found that individuals with MLD differed significantly from TD children, and concluded that it is not the specificity, but the severity of the ToM deficit that is unique to autism. However, their data say nothing about the severity of the deficit in individual children, but rather addresses the pervasiveness of ToM task failure in children with autism, MLD and TD children. It therefore remains unclear whether ToM deficits differ in their degree of severity between children with ASD and those with MLD. Studies 2 and 3 of this thesis use a range of ToM tasks to look at the severity of any ToM deficit in children with MLD and ASD who fail FB tasks, whilst study 5 uses teacher ratings of everyday behaviour to examine the same issue. The specificity hypothesis will also be tested in Study 1, where performance on 3 different FB tasks and the relationship of that performance to language will be assessed in a large sample of children with MLD or ASD.

It would be a mistake to equate failing a FB task with the lack of a ToM, and in particular with deficits in an innate 'ToM mechanism', or an inability to form M-representations. Children may fail these tests for a variety of reasons. Developmentally, for example, a lack of communicative input might delay the performance of deaf children, since it appears that there is no fundamental reason why being deaf should impair a child's ToM, and indeed deaf children of deaf signing parents are not delayed. (Peterson & Siegal, 1995; Woolfe et al., 2002). Online, difficulties in flexibility and set shifting, or in language comprehension, may account for the poor performance of some children with MLD on the tasks. Wellman et al. (2001) suggest that whilst such problems cannot account for the leap in performance on ToM tasks seen between the ages of 3 and 4 in typical development, they may account for individual differences in performance between those ages – so, for example, a child with very good cognitive flexibility may be able to pass FB tasks at 3 years 6 months, whilst a child with poorer flexibility might not pass until the age of 4 years. This may also be the case in children with MLD. In chapters 5 and 7 of this thesis, the nature of failure on ToM tasks in children with MLD and ASD is investigated. A theme throughout this thesis will be the comparison of children with MLD who fail FB tasks with children with ASD who fail, and the comparison of

children with MLD who pass FB tasks with those with ASD who pass. In this way, differences in the profile of the two groups can be identified, possible strategies investigated and the specificity of the ToM deficit tested.

The universality of the ToM deficit within ASD is also disputed. A proportion of children with autism pass first order ToM tasks – the proportion varies across studies, but ranges between 15 and 55% (e.g. Baron-Cohen et al., 1985; Frith, Happé, & Siddons, 1994; Happé, 1994; Ozonoff et al., 1991). Some individuals with Asperger's Syndrome or high functioning autism even pass more advanced 2<sup>nd</sup> order ToM tasks (Bowler, 1992; Ozonoff et al., 1991). However, they continue to show many of the social and communication problems associated with ASD. How, then, can it be argued that ToM is a fundamental deficit in ASD, and that many of the diagnostic features of autism in fact stem from this deficit?

One possibility is that the ToM task success of some individuals with ASD is due to chance or to simple strategies which do not reflect any true social insight (Frith, Morton, & Leslie, 1991). Alternatively, it is possible that some individuals have genuinely better social insight, and that their superior task performance reflects this (Frith et al., 1994; Happé, 1993, 1994). It is the case that children with ASD who pass FB tasks do, in some cases at least, appear to differ from those who do not. For example, Happé (1995) showed that children with autism who passed first order ToM tasks had a much higher verbal ability than TD children or children with MLD. Frith, Happé and Siddons (1994) found that some children who passed FB tasks also performed significantly better on teacher ratings of everyday behaviour thought to require a ToM, whilst Happé (1994) found that children with ASD who passed FB tasks were better able to attribute mental states to story characters in short vignettes than were children who failed. There is some debate over how children with ASD who pass FB tests, perform well on other tests of ToM, and even demonstrate better real life performance on ratings of ToM competence in everyday life may be achieving this. The existence of such individuals could be taken to suggest that ToM is not, after all, a fundamental deficit in ASD, or that it is possible to develop compensatory mechanisms for dealing with (innate?) problems in ToM. Tager-Flusberg (2002) suggests that language could play a vital role here in allowing some able children with ASD to construct an alternative route to the understanding of minds. In this thesis, Study 1 (Chapter 3) investigates the difference in language ability between those who pass and fail FB tasks both in ASD and MLD, whilst Study

4 (Chapter 8) investigates the performance of those who pass ToM tasks, with the aim of identifying strategies that children with autism may be using to compensate for their difficulties. Study 5 (Chapter 9) looks at real life ratings of both ToM and executive functioning and relates these to performance on experimental tasks, comparing the ratings given to children with ASD and those with MLD (both FB 'passers' and 'failers').

The first issue that will be addressed in this thesis is the relationship between language and performance on ToM tasks in children with ASD and MLD. Study 1 will also investigate the specificity of a ToM deficit to autistic spectrum disorders, as compared to individuals with moderate learning difficulties.

## **Summary**

There is now a great deal of evidence indicating that individuals with ASD have difficulties in attributing mental states, both in experimental tasks and in everyday life. It is not clear how this deficit in ToM might relate to other difficulties seen in ASD. Nor is it clear whether this ToM deficit is in fact specific to individuals with ASD, or whether it might be shared with some individuals with MLD. Chapter 2 of this thesis will review the evidence for a relationship between language and ToM in children with ASD and MLD. Chapter 3 will present a study investigating this relationship and the specificity of a ToM deficit to individuals with ASD. Chapters 4 to 7 will discuss and present studies testing the relationship between ToM and EF in children with ASD and MLD, and the possible impact of intervention in either of these domains. In chapter 8, the performance of those able individuals with ASD who pass basic ToM tasks will be investigated, in particular in the context of possible verbal strategies. In chapter 9, teacher ratings of ToM and EF in everyday life will be used to compare both those with ASD and MLD, and also those who pass ToM tasks and those who fail, and finally, in chapter 10, the results from the whole thesis will be discussed and a theoretical perspective will be offered on their interpretation.

## **CHAPTER 2**

# **THE RELATIONSHIP BETWEEN THEORY OF MIND AND LANGUAGE**

Language is a crucial part of human social existence. We use language to communicate our needs, thoughts and feelings, to share experiences and to express opinions on the world at large. We talk to others, to ourselves, and to those who have no chance of understanding, such as animals or even inanimate objects. Those without language are severely disabled in their ability to interact and participate in the social environment. But just how important is language in the development of socio-cognitive understanding? Is language simply the medium through which we can express understanding of concepts such as a ToM, or is it a fundamental part of the construction of these concepts? Alternatively, does some sort of social understanding have to precede normal linguistic development, providing the impetus and context for language learning?

Theory of mind (ToM), and its importance in the conceptualisation of autistic spectrum disorders, has been discussed in detail in the previous chapter. To summarise briefly, a ToM is the ability to attribute thoughts and beliefs to people in order to predict and explain behaviour.

The concept of language needs less explanation. However, language has a number of components which may be assessed in a range of ways, including standard tests of receptive vocabulary or grammar, more specific measures of syntax or pragmatics, naturalistic measures based on free speech, or measures of productive, narrative language. Many studies use the British Picture Vocabulary Scale (BPVS; (Dunn, Dunn, Whetton, & Burley, 1999) or its American equivalent, the Peabody Picture Vocabulary Test (PPVT; (Dunn & Dunn, 1981). These are tests of receptive vocabulary, where the child is asked to point to a picture that matches a word spoken by the examiner. Other tasks assess syntax (e.g. the Test for Reception of Grammar (TROG), Bishop, 1989), or narrative language (e.g. The Bus Story, Renfrew, 1991).

This chapter reviews the research on the relationship between ToM and language to date. It will discuss the relationship in normal development but then

focus on ASD and the possibility of a special relationship in autistic spectrum disorders. Chapter 3 will then look at the relationship experimentally in children with ASD and MLD. The thesis will return to this theme in Chapter 8, where the role of language in those individuals with ASD who pass basic ToM tasks will be investigated.

### **What is the evidence for a relationship between ToM and language?**

Table 2.1 summarizes a number of recent studies that have looked at the relationship between language and ToM both in typically developing populations and in those with developmental delay or disorders. In order to include as many relevant studies as possible, a range of measures of social cognition have been included. These include not only FB tasks, but also tests of affective perspective taking, social problem solving and emotion recognition. The finding of an association between language and social cognition is remarkably robust. The only group study (as opposed to case studies) that has found no association was Baron-Cohen, Leslie and Frith (1985), who found no effect of VMA on FB performance in an autism group. All the other studies found a significant correlation between language ability and socio-cognitive skills, despite using a wide variety of measures and samples. Some found specific rather than general associations, for example, Tager-Flusberg and Sullivan (1994) found no association between a vocabulary test and ToM, but a significant association between syntax and ToM, and Dunn et al (1991) found an association between feeling state and causal talk and later ToM performance, but no association with total amount of talk or mean length of utterance.

*Table 2.1. Summary of studies on the relationship between social cognition and language*

<b>Studies with typically developing children only</b>			
<b>Authors</b>	<b>Participants</b>	<b>Measures</b>	<b>Relevant findings</b>
Dunn, Brown, Slomkowski, Tesla, & Youngblade (1991)	50 preschoolers, T1 CA = 2:9 T2 CA = 3:4	T1 – observations of family carrying out normal routine. T2 – social cognition tasks, FB and affective perspective taking tasks.	No relationship between total number of talk turns at T1 and socio-cognitive measures at T2. Child's mean length utterance at T1 associated with affective perspective taking tasks but not with FB at T2. Feeling state and causal talk at T1 associated with performance at T2.
Jenkins & Astington (1996)	68 preschoolers. CA: 2:11 – 5:5.	4 FB tasks. TELD. Verbal memory	FB significantly associated with general language ability and verbal memory with age partialled out.

Authors	Participants	Measures	Relevant findings
Charman & Shmueli-Goetz (1998)	40 children CA: 6:11 – 8:10	2 <sup>nd</sup> order FB, belief-reasoning task CELF-R Narrative language measure	Strong relationship between CELF-R and FB, no relationship between narrative language and FB.
Hughes (1998a)	50 preschoolers CA: 3:3-4:7.	4 FB tasks, 2 deception tasks. BPVS and Bus Story (narrative language).	Language correlated significantly with all of the ToM tasks with age partialled out.
Astington & Jenkins (1999)	59 preschoolers in 3 consecutive years. T1 CA: 2:9 – 3:10.	2 FB, A-R. TELD	Language at T1 predicted ToM at T2 and T3. Not true vice versa. Syntax subscore particularly important – made an independent contribution to prediction of change in ToM scores.
Cutting & Dunn (1999)	128 urban preschoolers. CA: 3:6-4:10.	8 FB tasks. Emotion understanding tasks BPVS and narrative task.	FB, EU and language highly correlated, remained significant after age was partialled out.
Hughes & Cutting (1999)	238 same-sex twins (i.e. 119 twin pairs). CA: 3:7-3:9	8 FB, 2 deception tasks Verbal IQ	ToM and VIQ correlated significantly.
Watson, Nixon, Wilson, & Capage (1999)	Study 1 26 preschoolers CA: 3:9 – 6:8. Study 2 52 preschoolers CA: 3:4 – 6:11.	2 FB tasks. Test of Auditory Comprehension of Language (TACL-R). Social skills – teacher ratings.	Study 1: Positive but non-significant association between FB and language. Study 2: Language comprehension was significantly correlated with FB and social skills score.
De Villiers & Pyers (2002)	28 preschoolers CA: 3:1 – 3:10 at start of study, Seen 4 times over the course of one year	3 FB tasks (including justification and explanation) Memory of complements, spontaneous speech, understanding of embedded questions	Memory for complements predicted later FB when performance on all aspects of tasks taken into account and not vice versa. With just FB prediction, memory for complement predicted FB and vice versa.
Hale & Tager-Flusberg (2002)	60 preschoolers CA 3:0 – 4:10 Divided into 3 training groups	Trained children on either FB (with no mental state terms), sentential complements (no mental state terms) or relative clauses.	Significant improvement on FB in FB and sentential complement trained groups. FB trained group did not improve on sentential complements, Relative clauses trained group did not improve on either complements or FB.
Lohmann & Tomasello (2002)	138 preschoolers CA: 3:3-3:10. Divided into 4 training groups.	Trained children in A-R with deceptive objects using varying degrees of language. Tested FB before and after.	Training using deceptive objects and either sentential complements or discourse or in complements alone. => FB improved. Exposure to deceptive objects with no language did not improve FB.
<b>Studies with individuals with developmental or acquired disorders</b>			
Authors	Participants	Measures	Relevant findings
Baron-Cohen, Leslie, & Frith (1985)	20 children with autism. CA: 6:1 – 16:6 14 children with Down Syndrome CA: 6:3 – 17:0 27 TD children.	2 FB. BPVS	CWA most likely to fail the FB task, despite advantage of VMA and NVMA. CWA who passed both FB trials no different in VMA or NVMA to those who failed.

Authors	Participants	Measures	Relevant findings
Ozonoff, Pennington, & Rogers (1991)	23 individuals with ASD CA: 8:0-20 20 clinical control group, with diagnoses including dyslexia, ADHD and learning disabilities. CA: 8:0-19	ToM battery, including 1 <sup>st</sup> and 2 <sup>nd</sup> order FB, A-R. Emotion perception task. Verbal learning and memory task WISC-R or WAIS-R	VIQ highly correlated with composite ToM score for both groups, correlation higher in the ASD group. VIQ was associated with the emotion perception score for the control group but not for the ASD group.
Tager-Flusberg & Sullivan (1994)	28 individuals with ASD or PDD. CA: 6:11-22:2 28 individuals with MLD CA: 7:5-17:7 18 ND children CA: 7:10-10:3	4 FB Explanation of action task – 9 ‘mental state’ stories and 3 control stories. PPVT-R and sentence structure subtest of the CELF.	VIQ correlated with FB and ‘explanation of action’ in both the ASD and MLD groups. CELF (syntax) correlated with all ToM tasks in ASD group. PPVT not correlated with ToM.
Sparrevohn & Howie (1995)	30 children with ASD in two groups. Lower VMA group mean CA = 9:0, Higher VMA group mean CA = 11:5.	4 1st order FB, 2nd order FB. PPVT Nonverbal ability test. Social behaviour rating scale.	CWA with a higher VMA (>7 yrs) did better on ToM tasks than did those with a lower VMA (<7 yrs).
Happé (1995)	70 young TD children CA: 3:1 – 4:9 34 children with MLD CA: 8:3-18:4 70 children with ASD CA: 6:3-18:7	2 FB tasks BPVS	BPVS and FB tasks correlated in MLD and ASD. ASD group required much higher VMA to pass FB, improvement in FB as language improved was more gradual than in the ND or MLD group.
Dahlgren & Trillingsgaard (1996)	20 children with ASD CA: 6:3 – 15:6 20 children with A.S CA: 7:3 – 12:9 20 normal children CA: 6:3 – 13:1 All groups were within normal range of intelligence.	1st and 2nd order FB tasks. Verbal IQ. Performance IQ.	Autistic or AS participants who passed both FB tasks had a higher VIQ than those who failed one or more tasks.
Gale, de Villiers, de Villiers, & Pyers (1996)	23 oral deaf children with normal non-verbal IQ CA: 4:0-9:0 Hearing, age-matched controls	Non-verbal FB 2 verbal FB tasks PPVT-R, verbal explanation of action.	Deaf children delayed both on verbal and nonverbal FB, requiring a CA of over 7 yrs to pass either type of FB task. Performance on the FB tasks closely related to language, particularly ability to produce complex sentences with cognitive state verbs.
Shields, Varley, Broks, & Simpson, (1996)	20 language impaired children (10 with a disorder of phonology and syntax, 10 with a disorder of semantics and pragmatics). 10 children with HFA. 10 TD children. CA for all: 7:0 – 9:0	FB, deception, Eye Direction Detection task, Social Comprehension questions.	ASD and semantic-pragmatic language disorder groups did worst on social cognition tests. The group with phonologic-syntactic deficits was worse than the control group, but better than the other clinical groups.

Authors	Participants	Measures	Relevant findings
Lord & Pickles, (1996)	51 children with ASD CA: 2:0 – 5:0 43 MLD/ language impaired CA: 2:0 – 5:0	Videotaped parental interviews, questionnaires measuring social and linguistic skills.	Children with language impairments showed nonverbal social communication deficits whether or not they had ASD.
Hadwin, Baron-Cohen, Howlin & Hill (1997)	30 children with ASD in 3 groups, trained in emotion, ToM and pretence. CA: 4:4 – 13:7	Trained children on either ToM, emotion or pretence over 8 days. Tested narrative skills before and after, and counted mental state terms.	No effect of any of the training on narrative skills or mental state terms.
Cohen, Menna, Barwick, Im, & Horodezky (2000)	380 children referred to child psychiatric services, some with language impairments. CA: 7.0-14.0 years	Social cognition – emotional decoding, social problem solving and self-cognition. 10 standardised language tests.	Children with language impairments had greater deficits in social cognitive processing than those with TD language.
De Villiers et al, (in prep) described in De Villiers (2000)	27 oral deaf children CA: 5:2 – 10:1 28 hearing preschoolers	6 non-verbal FB Verbal FB task Index of Productive Syntax (IPSyn) and PPVT.	All language measures correlated significantly with both verbal and non-verbal FB tasks. Performance on verbal and non-verbal FB tasks highly correlated.
Miller (2001)	10 children with SLI CA range 4:5-7:1 2 TD control groups (age-matched, other language-level matched)	Four FB tasks, varied as to linguistic complexity of task.	SLI group performed similarly to the age-matched on linguistically simpler FB, but more similarly to the younger group on those tasks with higher language demands.
Varley, Siegal, & Want (2001)	2 single case studies of men with agrammatical aphasia	Modified picture FB tasks	Participants passed FB despite severe grammar impairments.
Lundy (2002)	34 deaf children of hearing parents CA: 5:7-10:5	9 FB questions Language proficiency profile completed by teachers	Strong relationship between FB and language level. Also related to CA.

Abbreviations used: CWA = children with ASD. VMA = verbal mental age. NVMA = non-verbal mental age. BPVS = British Picture Vocabulary Scale. PPVT = Peabody Picture Vocabulary Scale. TELD = Test of Early Language Development. CELF - R – Clinical Evaluation of Language Fundamental – Revised. FB = false belief. A-R = Appearance-Reality. ToM = ToM. EU = emotion understanding. VIQ = verbal IQ. MLD = moderate learning difficulties. SLI = Specific Language Impairment, TD = typical development. CA = CA. T1 = Time 1, T2 = Time 2 etc.

It appears that despite a wide range of tasks used to assess both social cognition and language, the picture is overwhelmingly that of a robust relationship between language ability and aspects of social cognition (mostly ToM, but also some other areas), both in typical development and in developmental disorders such as ASD.

Language impairment appears to have a relationship with non-verbal social communicative deficits independently of ASD. Lord and Pickles (1996) found a strong relationship between expressive language and social deficits (as assessed by parents) in language-delayed participants aged between 3 and 5 years both with and without ASD. The children with language delay were not compensating for their lack of verbal ability with non-verbal communication, and the severity of their language delay was associated with the severity of their social deficits. This relationship was



also found by Cohen et al (1998) who looked at social cognitive processing (facial emotion decoding and social perspective taking) in psychiatrically referred children with and without language impairments.

Further evidence for a strong relationship between language and ToM comes from deaf children. These individuals make up a particularly interesting group because their exposure to language is often seriously reduced in the early years of life. Several studies have found that deaf children who come from non-native signing families show a tendency to be delayed in passing FB tasks (De Villiers, 2000; Lundy, 2002; Peterson & Siegal, 1995) in contrast to deaf children born to deaf parents, who are exposed to fluent signing from their earliest days. This is not simply due to the linguistic demands of the task, since an unpublished study described by de Villiers (2000) found that deaf children passed a non-verbal FB task on average at the age of 7 years 3 months, whilst hearing controls passed at 4 years 4 months. Peterson and Siegal (1995) suggest that this may be due to a lack of conversational input in the deaf children at a critical time in brain development, which means that even when the children get to school and meet signing peers and teachers, it takes them a long time to catch up on their social cognitive skills. De Villiers (2000) has a different perspective and argues that the delay is due to the late learning of complement syntax, a grammatical structure that she claims is crucial for ToM development. Her recent longitudinal study (De Villiers & Pyers, 2002) found that children's memory for complement syntax predicted later performance on a range of FB tasks (including explanation and justification tasks), whilst the converse was not true. However, if only the standard 'prediction' FB task was used, the direction of the relationship went both ways.

These findings with deaf children do seem to indicate a crucial role for language in early acquisition of ToM. However, their deafness may have other consequences – for example, it may be harder for a deaf child to establish joint attention with a hearing parent without auditory cues. This could result in delayed ToM development because the precursors are also delayed, due to the child's deafness. A longitudinal study of joint attention, ToM and language in young deaf children, both of deaf and hearing parents would help to clarify this issue.

## **How are language and ToM related?**

There are several possible explanations for the relationship between language and social cognition. In this chapter I will concentrate on ToM rather than addressing wider areas of social cognition such as emotion recognition, since much of the research on social cognition and language has centred around ToM, and it is the main area of this thesis. I will briefly describe several alternative explanations of the relationship, and will then discuss them in the light of the available evidence.

### **Language is a precursor to theory of mind**

Language may be a necessary precursor to ToM. There are two alternative possible pathways here. It is possible that language is necessary for ToM development because the child needs to engage in and hear conversations about the mind and mental states in order to learn about this area for themselves (Dunn et al, 1991). Alternatively, syntactic language may itself provide a representational 'bootstrap' for ToM – the structure of language in itself introduces the possibility of different points of view, and provides the structure for symbolic representations of the type necessary for ToM understanding (Tager-Flusberg, 2000, De Villiers, 2000). Of course, this second account does not claim that conversational input is unimportant, but that exposure to specific aspects of language is what a child needs to develop a ToM – therefore a child who was exposed to conversation but who never heard complement structures might be expected to have a ToM delay. However, this is a highly unrealistic scenario and in real life it is hard to pull these two aspects of language exposure apart.

### **Early understanding of mind is necessary for language development**

Whilst the ToM tested at age 4-5 by FB tasks is quite obviously not necessary for language development since most children are talking fluently by that age, early socio-cognitive skills such as joint attention could be a necessary precursor to language development. A child may have to have some understanding of the intentionality of other people in order to learn language (e.g. Baldwin & Saylor, 2002; Tomasello & Farrar, 1986).

### **Language and theory of mind develop separately and independently**

Language and ToM may develop separately and independently, and their association may be due to some other factor that relates to both. However, obvious

candidates such as age, family background and general ability can be partialled out without the association disappearing (Cutting & Dunn, 1999; Jenkins & Astington, 1996).

**Theory of mind tasks make demands on language ability which account for the relationship**

The association may be due to task-specific factors in ToM tasks such as demands on language ability (Siegal & Beattie, 1991).

I will start this discussion with the last option, as it is important to eliminate this possibility before discussing any more complex issues. Almost all these ToM tasks are linguistically mediated, and it is easy to imagine how young children might fail due simply to misunderstanding the requirements of the task. Factors other than language may also contribute to the failure of children below the age of 4 - for example executive function (abilities such as strategic planning, inhibitory control and working memory; (Russell, Jarrold, & Potel, 1994). These are discussed in detail in chapter 4. To control for these factors, many researchers have developed tests of ToM that attempt to eliminate or reduce the other (non-ToM) demands on the child. Several of these studies report above-chance performance from 3-year-olds, indicating that it may in fact be these other demands that account for their frequent failure (Siegal & Beattie, 1991). However, a meta-analysis of these studies, looking at 77 reports of 178 different studies found that despite these experimental manipulations, the findings overall are consistent with a conceptual change in social understanding occurring between the ages of 3 and 5 (Wellman et al., 2001). Non-verbal FB tasks correlate with language ability highly in deaf children (De Villiers, 2000), again indicating that it cannot be the language demands of the task alone accounting for the association. It seems, therefore, that we are justified in looking more closely at the association.

It seems unlikely, given the evidence, that language and ToM in fact develop completely independently. No convincing third factor has been put forwards as yet which accounts for the association, and there is a great deal of evidence which indicates that the two are associated in a range of groups with a range of measures. This is not to say that we can rule out this possibility, simply that the evidence as it stands does not support this view.

If we take a narrow definition of ‘theory of mind’ as the conceptual understanding of minds measured by FB tasks, then it is not possible that this is a precursor to language as children do not pass these tasks until the age of 4, by which time they are already talking fluently. However, it is likely that earlier forms of understanding of minds – for example, the understanding of intentionality demonstrated in joint attention – are necessary for normal language development. Carpenter, Nagell and Tomasello (1998) demonstrated in a longitudinal study that the amount of time infants aged between 9 and 15 months spent in joint attention engagement with their mothers and the degree to which mothers used language following their child’s attentional focus predicted variation in the infant’s earliest communication skills, both linguistically and gesturally. Sigman & Ruskin (1999) showed that initiating and responding to joint attention was related to level of language skills (both concurrently and a year later) in children with ASD and Down Syndrome. This relationship could go some way to explaining the association between ToM and language, as joint attention may be a precursor of both ToM and communication and language skills. To rule this out, we need longitudinal studies which can look at the relationship between ToM and language whilst partialling out joint attention. However, this may be problematic since joint attention and ToM are so closely linked that joint attention may be considered more of a building block for ToM than a tool in its development, which it presumably is in the case of language. Partialling out variation in joint attention may therefore result in partialling out a good deal of the variance in ToM. It may be informative to look at different aspects of language – as I will discuss later, syntax appears to be particularly strongly linked to ToM, and it is possible that other aspects of language such as vocabulary may be more closely linked with joint attention. It is certainly possible to envisage how joint attention may be necessary for vocabulary development, as a child needs to appreciate which object is being labelled in order to learn its name, and there is lots of research which indicates that this is indeed the case (e.g. Charman, Baron-Cohen, Swettenham et al., 2001; Loveland & Landry, 1986; Mundy, Sigman, & Kasari, 1990; Mundy & Stella, 2000). It is less easy to see how joint attention might be directly related to grammar, although it may well be indirectly linked through vocabulary.

Finally, I will consider the option that language is in some way causally related to ToM development. There is a growing body of evidence, from work with special groups (e.g. De Villiers, 2000), longitudinal studies (Astington & Jenkins,

1999; Dunn, Brown, Slomkowski, Tesla, & et al., 1991) and most recently two training studies (Hale & Tager-Flusberg, 2002; Lohmann & Tomasello, 2002) that language has a crucial role in the development of a ToM. Lohmann and Tomasello's training study (2002) 3- to 4-year-old TD German children, trained large numbers of children (who initially failed at least one FB task) on perspective taking using a range of deceptive objects. There were five training conditions. The basic training revolved around children being shown objects first deceptively and then with the true nature of the object revealed – for example they were shown a rose that was in fact a pen. A third person then entered and the child was asked what he would think the object was. In the 'full training' condition this was discussed with the child using sentential complements, in half the cases using mental state verbs (24 children), in the other half using only communication verbs (24 children). In the Discourse only condition (30 children) no mental state terms or embedded complements were used in discussing the objects, but the child was still shown the deceptive objects and asked the questions. In the Sentential complements only condition (30 children) no deceptive objects were shown, the child was just exposed to complements using mental state terms. Finally, in the no language condition (30 children) the children were shown the deceptive objects with no language beyond 'look'. Children's performance on FB tasks (both an appearance reality task and FB transfer tasks) improved in all cases except for the no language condition. Those in the full training condition improved most, and it did not matter whether mental state verbs or just communication verbs were used. Those in the 'Discourse only' and 'Sentential complements only' groups both improved significantly. This study appears to indicate that without linguistic experience, children were not able to benefit from exposure to FB and deceptive situations. All the aspects of language varied appeared to contribute – and sharing perspective and labelling perspective appeared to have an additive effect. The improvement did not appear to be due to exposure specifically to mental state verbs, as communication verbs alone also had a similar effect on FB performance. Hale and Tager-Flusberg (2002) also found that training children on sentential complements improved their performance on FB tasks, even when no mental states were included in the training. They also trained children on FB tasks using a straightforward feedback and explanation paradigm (with no reference to mental states), and found that this group improved on FB tasks, but not on tests of sentential complements, suggesting that sentential complements are not a necessary part of passing FB tasks. A third training

group, in which the children were trained on relative clauses, showed no improvement on either sentential complementation or FB tasks.

The work with deaf children (De Villiers, 2000) has been described above. Deaf children of hearing parents (who have limited exposure to any sort of language for the first few years of life) show a delay in FB of around 3 years. It seems unlikely this is caused by their deafness since deaf children brought up in a signing environment from birth do not show any delay. This points to the importance of language as a necessary precursor to ToM, but does not inform us as to whether conversational aspects of language or syntax are more important, since deaf children have restricted access to both.

Both Astington and Jenkins (1999) and Dunn et al. (1991) found that early language predicted later ToM performance in young TD children. Their language measures were quite different – Astington and Jenkins used the CELF-R, and found that syntax was particularly strongly related to ToM. Dunn et al used more naturalistic measures and found that feeling state and causal talk was related to later ToM.

Given that the evidence for language being important for the development of ToM appears quite compelling, it would be useful to examine more closely the possible mechanism of this relationship. As stated above, there are two suggested alternative routes. Firstly, language may provide children with access to the social world, providing children with experiences of mental state talk and the opportunity to learn about the mental states of others. This is supported by Dunn et al (1991) who found that the amount of feeling state and causal talk in young children's families at age 33 months was related to their performance on FB tasks seven months later. Dunn and Brophy (2002) suggest that discourse and conversation are crucially important for a child's developing understanding of mind, and that the properties of specific relationships that the child has (e.g. with their parents or with their peers) predict individual differences in later ToM. In this case, presumably general rather than specific aspects of language should predict ToM, since those children with better general language skills would have superior access to the social environment. There is evidence that verbal ability as measured by standard tests such as the BPVS is related to ToM in normal groups (Cutting and Dunn, 1999) and in ASD (Happé, 1995). However, the problem with these findings is that they do not inform us as to the causal direction of the relationship, and it is plausible that those children with a

good ToM (or good precursors to a ToM) are able to acquire a better vocabulary because of their superior ability to infer a speaker's intentions. Even in the Dunn et al (1991) study, it may be that at the earlier time point some of the children already had superior socio-cognitive skills that stimulated greater amounts of feeling state and causal talk within the family. The work with deaf children could also be interpreted as supporting this more general account, as deaf children who are delayed in acquiring language have less access to the social world, and less experience of conversations and interactions (Peterson & Siegal, 1995). However, as discussed above, alongside their lack of experience of interaction goes a lack of exposure to syntax, and there is also a possibility that their deafness causes early difficulties in establishing joint attention with a hearing parent. Lohmann and Tomasello's (2002) training study also appears to suggest that many aspects of language contribute to FB understanding, and it may be the experience of deceptive situations in conjunction with some sort of language describing and contrasting them that allows children to gain a representational understanding of mind. However, it is striking that training children on sentential complements alone (with no deceptive situations) also improved their performance on FB tasks, although inevitably this training involved some reference to points of view.

The alternative option is that language itself provides the structure for understanding representation, and therefore is a bootstrap to ToM. Gaining the concept of a ToM requires symbolic thought, and language makes this a practical possibility. In non-social domains, there is evidence that providing a structure through language can aid children in making comparisons – for example in cases of relative size (Baldwin & Saylor, 2002) and also that labelling a dimension of a multi-dimensional object makes flexibility between other dimensions more likely in 4-year-olds (Zelazo & Jacques, 2002). It is therefore possible that the use of language may enable children to make comparisons between mental states that they would otherwise not have been able to do.

Syntax appears to be particularly strongly related to ToM - Astington and Jenkins (1999) in a longitudinal study of young TD children, found that the syntax subscore of their language measure was the only language subscore making an independent contribution to predicting later ToM scores. Tager-Flusberg and Sullivan (1994) also found that a syntactic language measure correlated highly with ToM performance in participants with ASD, whilst a vocabulary test was not associated at

all. Whilst it is possible to envisage how a better ToM could lead to an enhanced vocabulary because of the social learning of words (Bloom, 2000), it is harder to envisage how a better ToM could lead to better syntactic knowledge. Therefore, an association between syntax and ToM is more likely to mean that a sophisticated syntactic knowledge is a precursor of a ToM than vice versa. This syntactic knowledge may provide children with the structure to consider propositional attitudes (i.e. an attitude which may or may not be true). De Villiers (e.g. De Villiers, 2000) suggests that complement syntax in particular provides the bootstrap for comparing points of view and therefore gaining a ToM. For example, in a sentence such as ‘Sam thinks that Anne has a dog’, the complement is ‘Anne has a dog’, and it has a point of view marker ‘that’. The complement may be true or false, without affecting the truth value of the whole phrase – in other words, Anne may or may not have a dog, but it can still be true that Sam *thinks* that she has one. This is not the case with verbs such as ‘want’ – if we say ‘Sam wants Anne to have a dog’, there is no point of view complement and ‘Anne to have a dog’ does not have separate truth value to the truth of the whole phrase. De Villiers postulates that this distinction between the truth value of the complement and the truth value of the whole phrase is the crucial element that enables children to consider different perspectives and thus gain a ToM. However, since in English the only verbs that take a tensed complement are mental state and communication verbs, it is hard to tease apart the causal direction – use of these complements may indicate the existence of a ToM, rather than being a precursor to it. In contrast, in German the non-representational verb ‘want’ takes a tensed complement, and Perner and Zauner (2002) have shown that young children understand and use it well before they use representational verbs such as ‘think that’ or pass FB tasks. However, De Villiers counters this by pointing out that even though verbs such as ‘want’ take a tensed complement in German, this complement does not have the potential for different truth values that it would in the case of a verb such as ‘think’ and therefore does not provide a fair comparison (De Villiers, 2002). Against De Villiers’ claim, Varley et al (2001) present two case studies of adult men with agrammatical acquired aphasia. Despite serious grammar impairments, both of these men were able to pass FB tasks. Since these men presumably did have the necessary grammatical structures at earlier stages of their life, we might argue that the relationship is a developmental one, and once the ability to attribute mental states has been acquired it is not necessary to retain the ability to use syntax to retain the



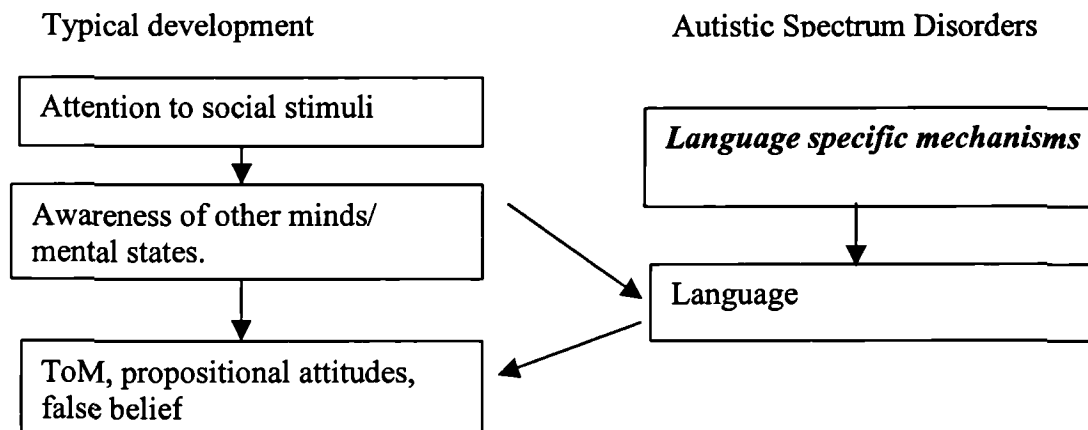
concept. Since these are single cases we should be careful about drawing too many conclusions, even about the online relationship – a way of investigating this further might be to temporarily selectively lesion brain areas relating to grammar using TMS, and to test the participants on FB tasks.

From the available evidence, it appears that in the early stages of language acquisition joint attention (and therefore early social understanding) is important (Carpenter, Nagell, & Tomasello, 1998; Sigman & Ruskin, 1999). This early understanding of intentionality may be particularly important in the development of vocabulary. Through this the child is able to acquire language, and through an increased competence in syntax and improved understanding of the social world around them is then able to form a concept of mental states and a ToM. It is very hard to pull apart the contributions of syntax and social interactions, as there are no real life cases of people who are exposed to syntax but no social interactions, or social interactions but no syntax. Lohmann and Tomasello's (2002) training study suggests that both discourse and complement syntax may be independently important. Problems at any stage of this development could lead to difficulties in ToM – we know that children with language impairment show socio-cognitive deficits (Lord & Pickles, 1996) as do deaf children (Peterson & Siegal, 1995), and it appears that children with ASD show less joint attention behaviour than their TD peers (Sigman & Ruskin, 1999).

Most of my discussion of the mechanism of the relationship between ToM and language so far has focused on TD children. ASD provides a special case of individuals with impairments in both ToM and language, and also a case where the relationship between the two areas may be quite different to that seen in TD children. There is evidence that both vocabulary (e.g. Happé, 1995) and syntax (e.g. Tager-Flusberg & Sullivan, 1994) are related to FB performance in ASD, and there is some evidence that children with ASD require a higher level of language ability than TD children in order to pass FB tasks. Happé (1995) suggests that a threshold effect may exist. In her sample, children with autism with a VMA of less than 5 years 6 months invariably failed FB, and those with a VMA over 11 years 2 months all passed. Given that the tasks themselves do not require such a level of language ability (as they are normally passed by 4-year-olds), what might it be about children with autism that means that they require such a relatively high level of language? Tager-Flusberg (2002) suggests a model of ToM development in which both language and mental

state awareness contribute in normal development. She postulates that in ASD an awareness of mental states (and possibly the attention to social stimuli) is lacking, and therefore language is the only potential route to ToM. Language ability therefore becomes crucially important in the development (or not) of ToM, having to compensate for lack of other sources of learning about mental states such as pretend play and friendships.

**Figure 2.1 Tager-Flusberg's model of ToM and Language development (adapted from Tager-Flusberg, 2002)**



Tager-Flusberg (2002) suggests that children with ASD use logical reasoning to pass FB tasks, in a way that TD children do not have to, because they do not possess a 'language of thought', by which she means an intuitive understanding of mental states. She postulates that communication verbs may be of particular importance as they can provide children with ASD with concrete examples of people saying things that do not match reality, and therefore the children can gain access to differing points of view even without considering mental states. In this case, syntax in particular should be important in predicting ToM performance, and this relationship should be stronger in individuals with ASD than in language or age matched controls without ASD. As yet, the studies demonstrating this relationship have either been with fairly small samples (Tager-Flusberg & Sullivan, 1994) or have only used one language measure (Happé, 1995). The next chapter in this study therefore reports a large study of FB understanding and language in children with ASD and MLD. Receptive vocabulary and grammar measures were used, along with three FB tasks.

## **Summary**

This chapter reviewed available research on the relationship between language

and ToM, and considered possible explanations of the relationship. From the available evidence it seems unlikely that the relationship is due to task demands alone, or that both develop separately and independently with the association being due to some third factor (although joint attention is a possibility). Theory of mind as measured by FB tasks cannot be necessary for language development, but earlier elements of ToM such as M-representations can, as can precursors such as joint attention. There is some evidence that indicates that language may be necessary for the development of a ToM, although the mechanisms behind this relationship are still debated. Language may be necessary as the route of access to the social world, to discussions about mental states and emotions, or syntactic language may provide the structure to enable the consideration of representations. Language may be particularly important in ToM development in children with ASD, providing them with an alternative route to some degree of social understanding. The following study examines the relationship between FB performance, vocabulary and grammar, in an attempt to clarify the causal relationship between ToM and language in ASD and MLD.

### **CHAPTER 3**

## **STUDY 1: HOW DO THEY PASS AND WHY DO THEY FAIL? THEORY OF MIND AND LANGUAGE IN CHILDREN WITH ASD AND CHILDREN WITH MLD**

### **Introduction**

As discussed in the previous chapters, the specificity and origins of ToM deficits in autism have been much debated (e.g. Baron-Cohen, 1989, 2000; Russell, 1997a; Yirmiya et al., 1998). This study aimed to test the specificity of a ToM deficit in ASD by comparing a large sample of children with ASD with a sample of participants of MLD of a similar verbal ability level. It also aimed to look more closely at the relationship between language and ToM by including measures of both vocabulary and grammar. An association between vocabulary and ToM does not inform us as to the direction of the relationship, since, whilst some degree of vocabulary may be necessary for the acquisition of a ToM, it is equally possible that possessing a ToM allows the acquisition of further vocabulary. Learning words from other people may require some understanding of their intentions and mental states, in order to deduce what they are referring to and therefore what the words mean. However, it is less likely that a ToM is required for the acquisition of grammatical structures, except in an indirect way through vocabulary. Therefore a relationship between grammar and ToM over and above that seen with vocabulary would suggest a causal role for language in ToM development in ASD. By comparing and contrasting the two groups, this study aimed to identify differences in the relationship between language and ToM in ASD and MLD, the hypothesis being that language (and particularly grammar) may play a more important role in ASD as it can provide an alternative route to ToM understanding, bypassing the awareness of other minds which non-autistic children (including those with MLD) have more intuitively (Tager-Flusberg, 2000).

The three main predictions for this study, therefore, were as follows. Firstly, children with ASD would be impaired on tasks of FB understanding when compared to those with MLD. Secondly, that language and FB performance would be particularly strongly related in ASD, and that grammar would predict FB performance over and above vocabulary, indicating a possible causal relationship from language to

ToM. Lastly, that FB performance in the MLD group would reflect task-specific factors rather than genuine difficulties in ToM, and that this would be reflected in a patchy pattern of task success and failure. Several subsidiary questions arose from these main predictions. These were: one, are the different FB tasks equally difficult? Two, are there gender differences in FB performance? Three, is a certain level of language necessary or sufficient for FB task success in ASD or MLD?

## **Method**

The participants in this study were 181 children and adolescents, recruited from schools for children with special educational needs in the greater London area. All children for whom parental consent was obtained and who were at school during the testing were included.

The autistic spectrum disorders (ASD) group contained 49 children who had a primary diagnosis of autism or an autistic spectrum disorder and 5 children with a diagnosis of Asperger's Syndrome. The remaining 9 (who were from a single school and who were all being educated in classes for children with autistic spectrum disorders) were described on their records as having social and communication disorders. These children were assigned to the ASD group after completion of a checklist of symptoms based on DSM-IV after discussion with their teachers.

The results are unchanged if these 9 children are excluded from the analysis. There were therefore 63 children in total in the ASD group. The age of the children in the ASD group ranged from 5 years 7 months to 16 years 2 months, with a mean of 10.74 years (s.d. = 2.35). There were 58 boys and 5 girls.

The remaining 118 children had non-autistic moderate learning difficulties and made up the MLD group. They were all being educated in schools for children with special educational needs, and were of mixed aetiologies. Children with fragile-X syndrome, or any suggestion of an ASD were excluded from this group. The age of the children in the MLD group ranged from 5 years 3 months to 14 years 5 months, with a mean of 12.13 years (s.d. = 1.75). There were 76 boys and 42 girls.

Table 3.1 gives age, gender and language characteristics of the two groups. Children whose scores fell below the lowest standardised or age equivalent score were credited with a value one unit below the floor value. This problem only applied to the BPVS standardised scores and the TROG VMA score. In the case of the BPVS std score the assigned value was 39, and applied to 3 children in the ASD group and 3

children in the MLD group. For the TROG VMA this value was 3.92 (i.e. 3 years 11 months). This applied to all children who passed fewer than 5 blocks, which was 13 children in the ASD group and 7 children in the MLD group. The large number of children affected by this floor effect was a reason for creating a TROG age-adjusted score that captured more of the variance in scores. All of the analyses were performed with both the age equivalent scores and the age-adjusted scores. Performance on the two language measures were significantly correlated (ASD VMA  $r = .73$  ( $p < .001$ ), standardised/age-adjusted  $r = .70$  ( $p < .001$ ) MLD VMA  $r = .62$  ( $p < .001$ ), standardised/age-adjusted  $r = .56$  ( $p < .001$ )).

**Table 3.1 Age, gender and language scores for the ASD and MLD groups; mean (s.d.) range**

	ASD (n = 63)	MLD (n = 118)
CA (years)	10.74 (2.35) 5.58-16.17	12.13 (1.75) 5.25-14.58
BPVS VMA	6.94 (3.26) 3.58-16.67	7.58 (1.95) 2.67-10.58
BPVS std	73.30 (19.20) 39-126	70.63 (10.82) 39-89
TROG VMA	6.30 (2.36) 3.92-11.00	6.16 (1.57) 3.92-10.00
TROG age-adjusted	88.37 (46.41) 0-198.24	91.47 (28.65) 13.80-159.47
Gender (boys/girls)	58/5	76/42

A one-way ANOVA showed no significant differences between the groups on any of the language measures. F-values ranged from 0.195 – 2.73, with all p-values above .10. There was a significant difference in CA ( $F=18.373$ ,  $df=179$ ,  $p < .001$ ).

### **Procedure**

All the children were tested individually by the researcher in a quiet room in their school. The testing session typically took 25 minutes. The tasks were presented in the following order: First Language measure, First FB task, Second FB task, Second Language measure. The order of both the language measures and the FB tasks was counterbalanced.

### **Measures**

British Picture Vocabulary Scale 2<sup>nd</sup> edition (BPVS II). Receptive vocabulary was assessed using the BPVS II (Dunn et al., 1999). In this test, children are shown four

pictures and asked to point to the picture which best tells the meaning of a word. The words become progressively harder. The test is discontinued when a child makes 8 errors in a group of 12. Raw scores were converted to a VMA and a verbal IQ score (std).

Test for Reception of Grammar (TROG). Receptive grammar was assessed using the TROG (Bishop, 1989). This task involves presenting the child with four pictures, and asking them to indicate which picture goes with a sentence containing a grammatical construct. Questions are arranged in blocks of 4, all of which test the same grammatical construct, and a child is said to have failed the block if they fail a single question. If the child fails 5 consecutive blocks the test is discontinued. Raw scores (number of blocks) were converted to a age-equivalent, referred to as a VMA, according to the manual. In addition, for comparison with the age-standardised score on the BPVS, an age-adjusted score was calculated as follows. The number of blocks passed was divided by the child's CA and this was then multiplied by 100. This gave a score that varied between 0 and 198 with a mean of 90.39 (s.d. = 35.74). This is referred to as TROG std.

#### False belief tasks.

Two different FB tasks were used, counterbalanced in order. One was an illustrated version of the standard unexpected transfer task (Wimmer & Perner, 1983). The script for the story is in Appendix I. The characters were a girl called Sally and a boy called David. Children were asked a FB question, a justification question and two control (memory and reality) questions. They were only credited with passing the FB question if they also passed both controls.

The second test was a standard deceptive box task using a Smarties tube. Children were asked both about another person's FB and about their own FB prior to the tube being opened. (Perner et al., 1987). The script for this task is shown in Appendix I. Children were asked two FB questions, concerning other's false belief and their own prior false belief, and a reality control question. They were only credited with passing the FB questions if they also passed the control.

There were therefore 3 FB questions asked, although only two FB tasks were carried out.

## Results

### How many children passed each FB question?

Table 3.2 gives the number and percentage of participants who either passed, failed, or failed the control question for each FB question. Two-tailed chi-square tests were used to compare the number of participants in each group who passed each task versus the number who failed. For the Sally-David task, the difference approached but did not reach significance ( $\chi^2=2.94$ ,  $df = 1$ ,  $p=.09$ ). For both Smarties questions the difference was highly significant (Self  $\chi^2=20.75$   $df = 1$ ,  $p<.001$ ; Other  $\chi^2=15.02$ ,  $df = 1$ ,  $p<.001$ ). Table 3.3 gives the performance on the 'Justification' question of the Sally-David task. Participants could only pass this if they had already passed the Sally-David FB question, and so only the results for those who passed that question are given.

***Table 3.2 Performance on the different FB tasks, split by group. Number (percentage) of participants who passed, failed or failed the control question.***

Task	ASD (n = 63)			MLD (n = 118)		
	pass	Fail	failed control question	pass	fail	failed control question
Sally-David (unexpected transfer)	25 (40%)	26 (41%)	12 (19%)	66 (56%)	38 (32%)	14 (12%)
Smarties other (deceptive box)	28 (44%)	29 (46%)	6 (10%)	90 (76%)	25 (21%)	3 (3%)
Smarties self (deceptive box)	31 (49%)	26 (41%)	6 (10%)	99 (84%)	16 (14%)	3 (3%)

***Table 3.3 Number (percentage) participants who passed the Sally-David justification question, having already passed the FB question.***

Task	ASD (n = 25)			MLD (n = 66)		
	Pass	incomplete/partial	fail	pass	incomplete/partial	fail
Sally-David justification question	16 (64 %)	3 (12%)	6 (24%)	46 (70%)	9 (14%)	10 (15%)

There was no difference between the groups ( $\chi^2 = 0.65$ ,  $df=1$ ,  $p = .72$ ).



### Participants who failed control questions

Those who failed control questions were considered to have failed to grasp the task, and therefore were not placed in either the 'pass' or 'fail' group as their ToM had effectively not been assessed in that task. Some children failed the control questions for only one task (8 in the ASD group, 17 in the MLD group), whilst others failed both (5 in the ASD group only). This group of five children who failed all the control questions had extremely low language ability (average BPVS VMA 3.10, average TROG VMA below floor) and were excluded from any further analysis. For those who failed only one control, their data was treated as missing for that task only.

When these five children had been excluded, the comparisons on age and language were repeated. Age was still significantly lower in the ASD group (ASD mean = 10.74, MLD mean = 12.13;  $F(1,174) = 19.17$ ,  $p < .001$ ). The BPVS standardised score was higher in the ASD group (ASD mean = 74.4, MLD mean = 69.8;  $F(1,174) = 3.89$ ,  $p = .05$ ). There were no significant group differences in the BPVS VMA score or either of the TROG measures. The difference in the BPVS standardised score reflected the lower CA but comparable VMA (as measured by the BPVS) of the ASD group.

### Did the FB tasks vary in difficulty?

Performance on the various tasks varied greatly, particularly in the MLD group, with over 85% passing the Smarties self question, but only 59% passing the Sally-David task. In order to look at the relative difficulty of the tasks, the number who performed inconsistently across the FB tasks were compared. If those individuals who perform inconsistently consistently fail one task rather than another, it would indicate that this task is harder overall for the group. Tables 3.4 and 3.5 gives the numbers of inconsistently performing participants who showed each pattern of passing and failing.

McNemar's test showed that there was no systematic difference in task difficulty for the ASD group (all  $p$ -values  $> .38$ ), but that in the MLD group there was a significant difference between the numbers passing Smarties self and other ( $p < .05$ ), between Sally-David and Smarties other ( $p < .01$ ) and between Sally-David and Smarties self ( $p < .001$ ).

**Table 3.4 Number of inconsistently performing participants with ASD who passed a particular task whilst failing another**

ASD	Passed		
Failed	Sally-David	Smarties-other	Smarties-self
Sally-David		8	13
Smarties-other	8		11
Smarties-self	8	8	

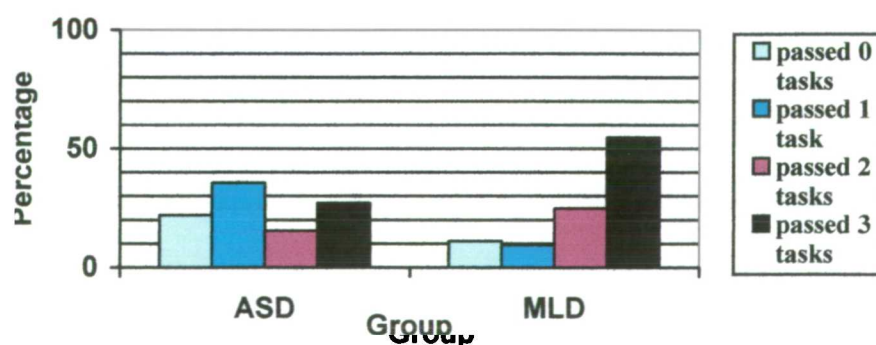
**Table 3.5 Number of inconsistently performing participants with MLD who passed a particular task whilst failing another**

MLD	Passed		
Failed	Sally-David	Smarties-other	Smarties-self
Sally-David		25	28
Smarties-other	7		12
Smarties-self	3	3	

#### How is FB performance distributed?

Figure 3.1 illustrates the percentage of each group who passed 0, 1, 2 or 3 of the FB questions in each group. Those who failed control questions were placed as follows: if they passed all the other FB questions they were given (and on which they passed controls), they were placed in the 3/3 group (this applied to 7 children in the MLD group and no children in the ASD group); if they passed one out of the two remaining FB questions they were placed in the 1/3 group (this applied to 4 children in the ASD group and 2 children in the MLD group); the remaining children failed all the tasks, even those they passed control questions on, and so they were placed in the 0/3 group.

**Figure 3.1. Percentage of each group who passed 0, 1, 2 or 3 FB question**



s

Performance does not appear to be dichotomous in either group, that is, children did not either pass or fail all the FB tasks. In the MLD group 65% of the group performed consistently, with 54% passing all tasks, and 11 % failing all tasks. In the

ASD group 48% of the group were consistent, with 22% failing and 26% passing all the tasks. A large proportion of both groups, therefore, showed inconsistent performance across the three FB tasks.

#### Characteristics of FB passers and failers

In order to ease comparison between the groups and with previous studies (e.g. Happé, 1995) a pass/fail variable was created. Children were credited with passing FB only if they passed all possible FB questions on which they had passed the control questions. For the majority of children this meant that they had to pass 3 questions. However, those children who failed a control question were scored as passing if they passed all the FB tasks on which they passed the control – i.e. either one or two questions. On this criterion, 15 of the ASD group (24%), and 65 (55%) of the MLD group were credited with passing FB. This difference was highly significant ( $\chi^2=16.29$ ,  $df=1$ ,  $p=.001$ ). If a less stringent criterion for ‘passers’ is adopted – i.e. passing 2/3 questions rather than 3/3, then 24 (38%) of the ASD group and 93 (79%) of the MLD group are ‘passers’. This difference is also highly significant ( $\chi^2 = 29.79$ ,  $df = 1$ ,  $p<.001$ ). Hereafter the more stringent criterion is used for passing all FB questions. The group characteristics of these ‘passers’ and ‘failers’ are given in Table 3.7.

Inspection of Table 3.6 suggests that the passers and failers in the ASD group differed on language ability more than did passers and failers in the MLD group: there was a difference of over 4 years on BPVS and TROG VMA in the ASD group in contrast to a difference of only about one year in the MLD group. A similar pattern emerges on the standardised measures. A one-way ANOVA showed that in the ASD group, passers and failers differed significantly on BPVS VMA ( $F(1,56) = 36.51$ ,  $p<.001$ ), BPVS std ( $F(1,56) = 29.54$ ,  $p<.001$ ), TROG VMA ( $F(1,56) = 116.75$ ,  $p<.001$ ) and the TROG std ( $F(1,56) = 46.46$ ,  $p<.001$ ). There was a trend towards a difference on CA ( $F(1,56) = 2.97$ ,  $p=.09$ ).

In the MLD group, a one-way ANOVA showed significant differences between the passers and failers in CA ( $F = 6.36$ ,  $df=1,116$ ,  $p<.05$ ), BPVS VMA ( $F=10.37$ ,  $df=1,116$ ,  $p<.01$ ), TROG VMA ( $F=10.88$ ,  $df=1,116$ ,  $p<.01$ ) and TROG std ( $F = 10.97$ ,  $df = 1,116$ ,  $p<.01$ ). The groups did not differ on the BPVS std score ( $F = 1.54$ ,  $df = 1,116$   $p=.22$ ).

**Table 3.6 Group characteristics of FB ‘passers’ and ‘failers’ mean (s.d) range**

	ASD ‘passers’ (n=15)	ASD ‘failers’ (n=43)	MLD ‘passers’ (n=65)	MLD ‘failers’ (n=53)
CA (years)	11.64 (1.44) 9.08-13.42	10.42 (2.61) 5.58-16.17	12.49* (1.27) 7.25-14.58	11.72* (2.13) 5.25-14.50
BPVS VMA	10.63*** (3.50) 4.92-16.67	6.10*** (2.06) 3.08-10.83	8.09** (1.64) 3.08-11.08	6.97** (2.13) 2.67-10.58
BPVS std	92.73*** (19.53) 48-126	67.98*** (13.44) 39-91	71.05 (10.61) 39-88	68.34 (13.05) 39-89
TROG VMA	9.30*** (2.01) 5.75-11.00	4.92*** (1.04) 3.92-9.00	6.49** (1.55) 4.00-10.00	5.59** (1.52) 3.92-10.00
TROG std	142.85*** (24.76) 100.00-198.24	78.02*** (33.72) 11.55-146.88	99.04** (22.14) 41.15-142.22	82.18** (32.92) 13.79-159.47

Significant differences between passers and failers within each diagnostic group are marked as follows:  
\*\*\*  $p < .001$ , \*\* $p < .01$ , \* $p < .05$

In order to get some idea of the magnitude of the differences, effect sizes ( $d$ ) were calculated for both groups. In the ASD group effect size ‘ $d$ ’ for differences between passers and failers ranged from 1.50 – 2.87 (with the TROG having higher effect sizes to the BPVS) indicating very large effect sizes (Cohen, 1969, suggests that 0.8 should be considered large). In the MLD group  $d$  ranges from 0.23-0.59, indicating low to moderate effects.

#### Are there gender effects on FB performance?

There has been some suggestion (e.g. Bosacki & Astington, 1999; Charman, Ruffman, & Clements, 2002) that girls may out-perform boys on ToM and verbal ability measures. The ASD group had only 5 girls in it, and thus numbers were too small to compare genders statistically, although the means suggested that the girls had a slightly superior BPVS VMA (8.57, versus 7.15 years for boys). On the TROG VMA the groups were very similar (girls mean = 6.30 years, boys mean = 6.03 years), as they were on the age-adjusted scores (BPVS std girls mean = 77, boys mean = 74; TROG std girls mean = 97, boys mean = 95). As far as FB performance goes, 2 (40%) of the girls with ASD and 13 (25%) of the boys were FB ‘passers’.

The MLD group had 42 girls and 76 boys. A one-way ANOVA showed that the two groups differed significantly on the language measures, with boys outperforming girls on the BPVS standardised ( $p < .01$ ) and VMA ( $p < .05$ ), and the

TROG age-adjusted score ( $p < .05$ ). The difference on the TROG VMA did not reach significance ( $p = .11$ ). A  $\chi^2$  analysis found no difference in number of boys and girls who passed FB. A forced entry logistic regression was done to test the effects of gender on FB performance in the MLD group, this showed that gender was not a significant predictor alone, nor was it an additional predictor once language was entered into the analysis.

### Language and FB

Table 3.7 shows the correlations between FB total score (numbers of questions correct, 0-3) and age and language characteristics in each participant group. When BPVS and TROG VMA were partialled out of the age x FB correlation this was no longer significant: ASD ( $r = .20$   $p = .13$ ), MLD ( $r = .06$ ,  $p = .53$ ). By contrast, FB x language correlations all remained highly significant when CA was controlled for.

Logistic regression analyses were performed to look for predictors of FB performance in the two groups. The two groups were examined separately because there is much prior evidence indicating that the relationship between language and FB is different in those with ASD and those with MLD. Of primary interest was whether the grammar measure (the TROG) predicted performance on FB over and above vocabulary (BPVS) and age. Therefore a forced-entry logistic regression was performed, with age and BPVS entered first, followed by TROG. In the ASD group in the first block age was entered with a non-significant improvement ( $\chi^2 = 2.92$ ,  $df=1$ ,  $p = .09$ ) In Block 2 BPVS VMA was then added with an improvement ( $\chi^2 = 22.17$   $df=1$ ,  $p < .001$ ), and was found to be a significant predictor ( $\beta = 0.69$ ,  $df=1$ ,  $p < .001$ ). In the final block TROG VMA was added with an improvement ( $\chi^2 = 17.58$ ,  $df=1$ ,  $p < .001$ ). In Block 3 TROG VMA alone was a significant predictor ( $\beta = 1.27$ ,  $df=1$ ,  $p = .01$ ). This final model using BPVS VMA and TROG VMA predicted 89.7% of cases (i.e. whether they passed or failed) correctly.

If the standardised scores are used the pattern is similar, except that CA is a significant predictor and remains so throughout, so that in the final model both the TROG ( $\beta = 0.10$ ,  $df=1$ ,  $p = .01$ ) and CA ( $\beta = 0.72$ ,  $df=1$ ,  $p < .05$ ) are significant predictors. This difference was expected since the TROG VMA and BPVS VMA are highly correlated with age. This final model predicted 91.4% of cases (pass/fail) correctly.

**Table 3.7 Correlations between FB, age and language for the ASD and MLD groups**

	ASD (n=58)	MLD (n=118)
FB x CA	.46**	.33**
FB x BPVS VMA	.67**	.50**
FB x BPVS std	.46**	.26**
FB x TROG VMA	.77**	.46**
FB x TROG std	.60**	.50**

\*\* p<.01

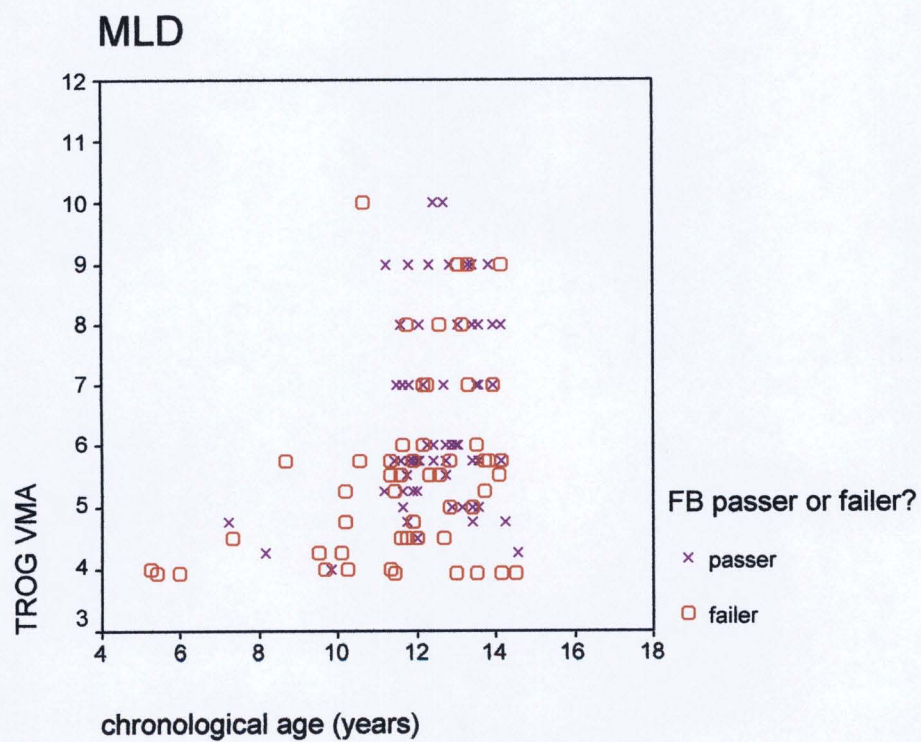
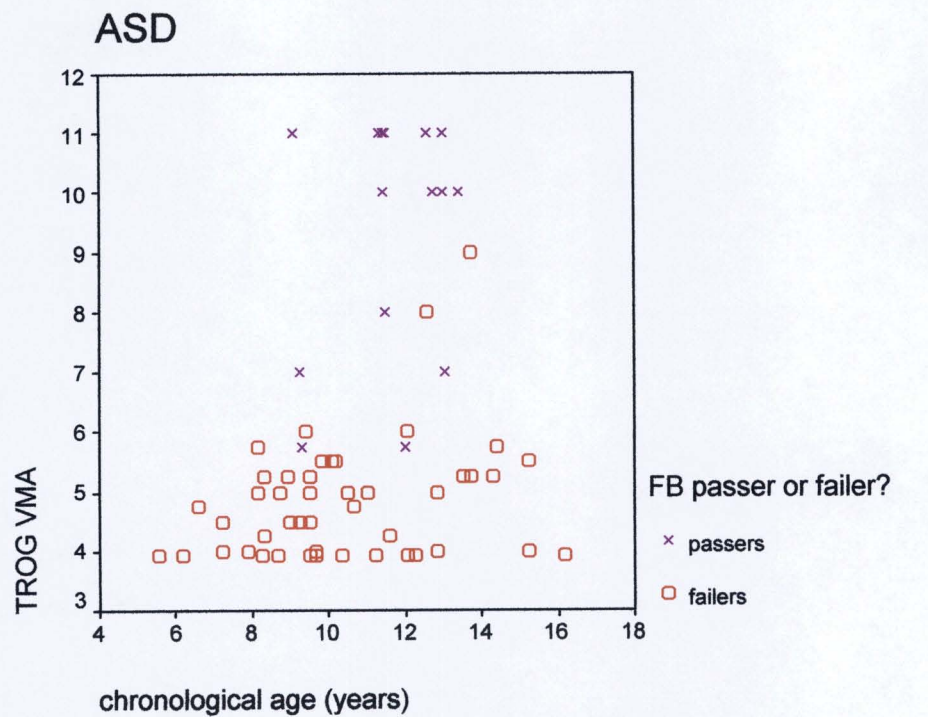
In the MLD group the same analysis was run. In Block 1, CA was added with significant improvement ( $\chi^2 = 5.98$ ,  $df=1$ ,  $p<.05$ ). In Block 2, BPVS VMA was added with an improvement ( $\chi^2 = 4.42$ ,  $df=1$ ,  $p<.05$ ). In the final block (when the TROG VMA was added) there was no significant improvement in the model ( $\chi^2 = 1.59$ ,  $df=1$ ,  $p= .21$ ). None of the 3 predictors were uniquely significant in this model (all p-values were greater than .22). In Block 2, the BPVS VMA alone was a significant predictor ( $\beta = 0.25$ ,  $df = 1$ ,  $p<.05$ ). This model using BPVS VMA only predicted 65% of cases correctly.

With the standardised scores the pattern is somewhat different in the MLD group. Adding in the BPVS std at Block 2 makes a non-significant difference ( $\chi^2 = 3.14$   $df=1$ ,  $p=.08$ ), and adding the TROG std at Block 3 makes a significant difference ( $\chi^2 = 4.76$   $df=1$ ,  $p<.05$ ). In Block 2 only CA is a significant predictor, and in Block 3 both CA and the TROG std are significant predictors. This difference may be due to the increased variance in the TROG std as compared to the TROG VMA (which shows a floor effect) and the reduced variance in the BPVS std (which also has a floor effect) as compared to the BPVS VMA. The best model in this regression (i.e. that in Block 3) predicted 62.7% of cases correctly, in contrast to the same analysis in the ASD group, which predicted 91.4% of cases correctly.

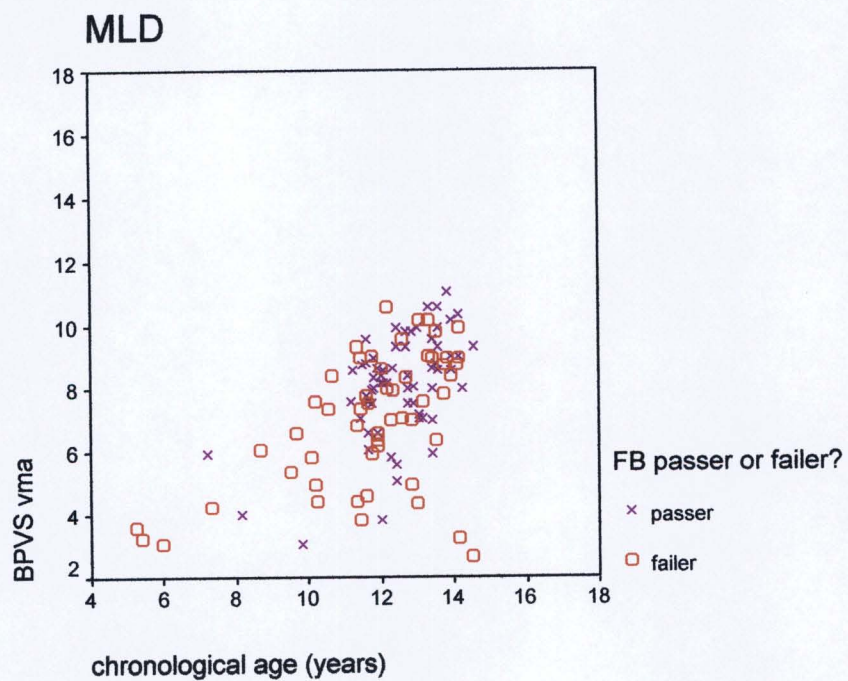
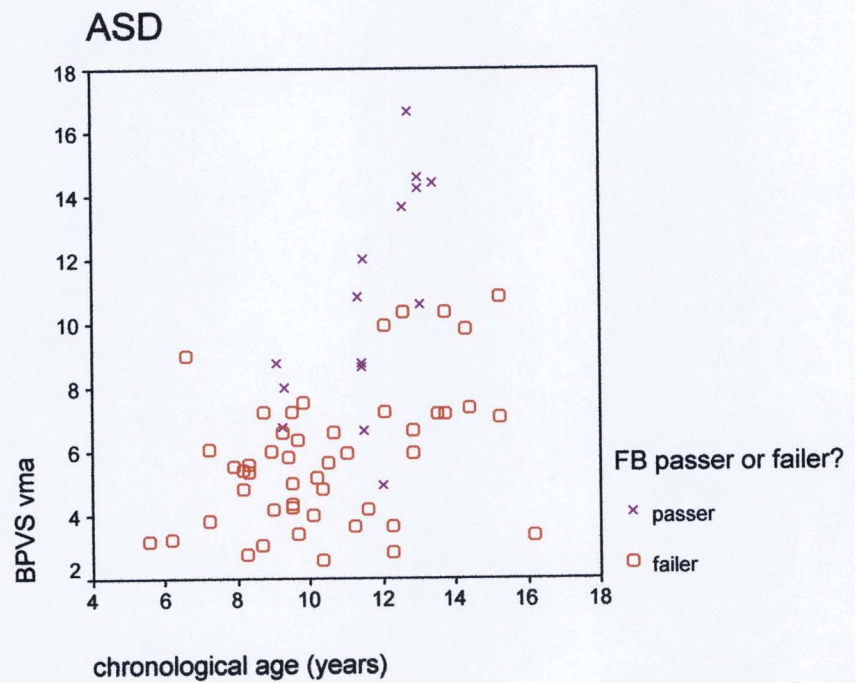
To examine the relationship between the two language measures and FB performance at a more individual level, scatter plots were inspected, showing TROG VMA against age, with markers set by whether the participant passed or failed FB. Figure 3.2 shows these plots. The VMA variables were chosen for this analysis because they are both conventional measures.



**Figure 3.2. Scatter plots of FB passers and failers on TROG VMA v CA.**



**Figure 3.3** Scatter plots of passers and failers on BPVS VMA v. CA





Inspection of figures 3.2 and 3.3 suggest that passers and failers in the ASD group were more easily distinguished by the TROG VMA than were passers and failers in the MLD group. In the ASD group no participant with a TROG VMA of less than 5.75 years (38 individuals) passed FB, whilst all those with a TROG VMA equal to or above 10 years (10 individuals) passed. A dividing line at a VMA of 6 years would divide passers and failers with the exception of four children. In the case of CA, no child younger than 9.08 years passed FB (14 individuals), but there is no upper threshold – the oldest participant in the study failed at the age of 16.17 years.

In the MLD group the relationship is far less clear. Individuals with TROG VMA scores of 4 years passed FB, and 20 individuals with scores less than 5.75 passed (the threshold in the ASD group). At the other end of the ability range there is no more clarity: one participant with a TROG VMA of 10.00 still fails FB. All children (3 individuals) with a CA of less than 7.25 years failed.

Figure 3.3 gives the equivalent graphs for the BPVS VMA. From figure 3.3 it can be seen that the BPVS distinguishes between passers and failers less well than the TROG, but it still divides the groups more clearly in the ASD group than in the MLD group.

In the ASD group all participants with a BPVS VMA of less than 4.92 fail (13 individuals). A single child passes with a BPVS VMA of 4.92, but no one else passes with a score of less than 6.67. All children with scores higher than 10 years 11 months pass (7 individuals).

In the MLD group again the pattern is less clear. One participant passes FB with a BPVS VMA of only 3.08 years. No children in the group had a BPVS VMA of more than 10.58, but one child with this VMA still failed FB.

## **Discussion**

This study aimed to investigate a number of questions. The three main predictions were that children with ASD would be impaired on FB as compared to children with MLD, that language (specially grammar) and FB would be particularly strongly related in ASD, and that task-specific factors would be important in determining performance on FB tasks in children with MLD.

### **False belief performance in ASD and MLD**

The prediction that children with ASD would perform less well on FB tasks than non-autistic children matched on VMA was confirmed, with 74% of the ASD group failing at least one FB task, whilst only 45% of the MLD group did.

The specificity of the ASD deficit in ToM is a topic of some debate, and some authors have argued that a deficit in FB performance can be seen in other groups with learning difficulties. (Yirmiya et al., 1998; Zelazo et al., 1996). This study found some support for that view. Since we did not have a control group of TD children, it is not possible to contrast our non-autistic group with normal development. However, it is striking that 45% of the MLD group failed at least one FB question, even though the majority of them had VMA scores of over 4 years (only 10 children in the group had either a TROG or BPVS VMA of less than 4 years), the age at which a TD child would be expected to pass. Even if we exclude all children with either a TROG or BPVS VMA of 5 years or less, (taking 5 as a conservative estimate of when a TD child would be expected to pass a FB test) 39% of the remaining MLD group (32 individuals) still failed. There are a number of possible explanations for their poor performance.

A first option is that many children with MLD do have genuine difficulties in ToM. Against this is that they do not appear to have the social and communication difficulties associated with this in ASD (although this has not been widely studied in individuals with MLD). Frith, Happé and Siddons (1994) found that in children with MLD FB performance did not relate to teacher ratings of real life social adaptation, implying that FB tasks may not be a good measure of true social cognitive ability in this group. This study will be discussed in greater detail in Chapters 8 and 9. Chapter 9 of this thesis will also investigate the relationship between FB task performance and real life behaviour in children with MLD and ASD. It is possible that these children are similar to TD 3-year-olds who fail FB tests despite demonstrating relatively competent social understanding through speech and behaviour (Bartsch & Wellman, 1995). However, those with MLD are obviously not equivalent to TD 3-year-olds in that they have language levels (and, presumably, general developmental levels) far above that of the average 3-year-old. It has been suggested that TD 3-year-olds may fail FB tasks due to executive difficulties (e.g. Roth & Leslie, 1998), and these could also be a source of problems for some of this group of children with MLD. More specifically, candidates for executive limitations of FB performance could include

inhibitory control or flexibility. This will be discussed in greater detail in chapters 4 and 5.

If some of the MLD group were failing these tasks because of task-specific demands rather than due to genuine ToM difficulties, we might expect some tasks to be notably easier than others. In contrast, if the ASD group was failing due to ToM problems, the task demands should be less significant and a similar percentage of the group should be failing each task. We therefore examined performance across tasks next.

### **Variation in performance across FB tasks**

Three different FB questions were included in this study (one memory of own FB, two prediction of other's FB), from two different tasks (Smarties and Sally-David). The two tasks included in this study varied in their task demands, with the Sally-David task having a more complicated narrative and being about fictional characters, as opposed to the Smarties task, which was simpler and about real people. There was a much greater variation in performance across FB tasks in the MLD group. Individually, the MLD group outperformed the ASD group on 2/3 of the questions. On the third question, (which was the Sally-David FB prediction task) performance was not significantly different. This was due to the MLD group performing particularly poorly on this task (as compared to the other FB tasks), rather than to the ASD group performing well. Amongst those in the MLD group who performed inconsistently across the three questions, the tasks clearly formed a hierarchy of difficulty, with 'memory for own FB' being easiest, the 'prediction of other's FB' in the Smarties task next, and the Sally-David FB prediction being hardest. In the ASD group there was no such hierarchy – if a child failed one task, they were as likely to fail the Sally-David task as they were to fail either part of the Smarties task. This pattern of performance perhaps suggests that task success in those in the ASD group who performed inconsistently was due to chance factors.

The relative difficulty of the Sally-David task for the MLD group could be for a number of reasons. This task was presented in an illustrated format in contrast to the Smarties task, which was about a real object. In addition, the Sally-David task was about fictional characters rather than real people (in the Smarties prediction task the children were asked 'If your friend (name) came in, what would he/she think was in the tube?'), and had no element of deception, whilst the set up of the Smarties task

made it more likely to appear like a trick, although deception was not explicitly referred to at any point. Deception is known to be a feature which makes FB tasks easier for young TD children (Wellman, Cross et al., 2001).

The difference in the MLD group between the ‘memory for own FB’ and ‘prediction of other’s FB’ questions in the Smarties task is particularly interesting as the task demands (e.g. EF) appear to be otherwise identical in this case. The ‘own FB’ question was always asked second, after the control question, but it might be expected that this would make it harder, as the child has just been asked again what is really in the tube and therefore that information has become more salient. It may be that in children with MLD there is a developmental progression in FB performance, with reflecting on one’s own FB being easier than predicting another person’s. It is also interesting that the ASD group showed no such progression, in their case the three tasks were equally difficult. This may indicate that children with ASD have a general difficulty in ToM that over-rides other task factors, whilst in children with MLD their problems in FB tasks are in many cases due to task-specific factors (e.g. executive or language demands of particular tasks). It is particularly striking how hard the MLD group found the Sally-David task in contrast to the Smarties ‘other’ task (56% passed Sally-David as opposed to 76% passing Smarties ‘other’) – questions which might have been assumed to be equivalent in their ToM demands. This finding is unexpected, as other studies of this type (e.g. Grant, Grayson, & Boucher, 2001) have found that performance across tasks such as these is largely consistent in children with ASD (77% of the group either passed or failed all tasks in the Grant et al study). However, Grant et al’s findings are in line with ours in that they found a lack of agreement between performance on a ‘Sally-Anne’ type task and a deceptive box type task, similar to the ‘Smarties other’ task used in this study. This was in fact the only comparison in the Grant et al study that did not meet their criterion for acceptable agreement (they had a ‘three boxes’ task and a second deceptive box task in which the child was asked what a toy hippo would think was in the box rather than what another child would think) indicating that these two tasks may be particularly different in the demands they make on children.

### **Gender differences**

Some recent studies have suggested that girls may show slightly superior performance on social cognitive tests as compared to boys (Bosacki, 2000; Charman

et al., 2002). In order to look at this issue in children with MLD and ASD, we compared scores on the FB tasks. Gender comparisons were only possible in the MLD group, as there were only 5 girls in the ASD group. The boys in this sample were more verbally able than the girls, an unexpected result, probably due to having drawn our sample from special schools, which have a higher proportion of boys generally and probably represent a wider range of ability in boys as compared to girls. This may be because boys are more likely to demonstrate disruptive behaviour than girls, and therefore low ability girls are less likely to cause problems in a mainstream classroom and to be sent to a special school. However, there was no gender difference in FB performance and gender did not predict performance on FB in this sample.

### **Language and FB**

The second major area of interest in this study was the relationship between language and FB performance in ASD and MLD. Happé (1995) found that performance on the BPVS predicted performance on FB tasks in children with ASD, but not in children with MLD. In contrast, Tager-Flusberg and Sullivan (1994) found in a much smaller sample (21 children with ASD with FB data as opposed to 70 in the Happé study) that performance on the American equivalent of the BPVS (the PPVT) did not correlate with FB performance, whilst performance on a grammatical test did in both ASD and non-autism groups. Their sample was unusual in that 90% of the ASD group passed at least three out of four FB questions. In the present study, FB performance correlated significantly with both BPVS and TROG in both groups.

How to standardise raw scores from measures such as the BPVS and TROG has long been problematic in samples of this nature. Most studies with children use VMA scores, which equate a child's performance with a developmental level. An alternative is to use age standardised scores, which give a measure of ability relative to children of the same age, and which should remain relatively stable over time. Therefore, a child of 5 who achieves a certain raw score on a task such as the BPVS may achieve a standardised score of 100, whilst a 10-year-old who achieves the same raw score would have a standardised score of 50. However, they would both be given the same VMA of 5 years. This can be problematic in that it assumes that a 5-year-old of average ability and a 10-year-old of subnormal ability are developmentally equivalent. All analyses in this study were done using both types of scoring,

however, the VMA is the most conventional for groups of children such as the ones in this sample and therefore is the one on which most of this discussion is based, and which is used through most of the rest of this thesis.

A hypothesis for this study, following Happé (1995), was that receptive vocabulary would predict FB performance in children with ASD, but that receptive grammar would also be a significant predictor over and above vocabulary. The prediction, following a range of studies (e.g. Capps, Kehres, & Sigman, 1998; Tager-Flusberg & Sullivan, 1995) was that the relationship between language and ToM would be stronger in children with ASD than in those with MLD.

This hypothesis was supported by the data. Vocabulary was a significant predictor of FB performance in children with ASD. However, grammar was the best predictor, and when it was entered into a forced-entry logistic regression the effects of vocabulary became non-significant. In contrast, in the MLD group the pattern was far less clear – logistic regression models including age, vocabulary and grammar did not fit particularly well, the best models predicting only 65% of cases correctly (in contrast to 92% of passers and failers in the ASD group). There was also no clear indication of whether grammar or vocabulary was most important – when the regression was done using age equivalents, vocabulary was the best predictor, whilst when it was done using age-standardised scores grammar was. When the relationship was investigated further, it appeared that in the ASD group there was a clear threshold effect for both the TROG and the BPVS VMA. Again, the relationship seen in the MLD group was far less predictable. Individuals passed and failed at every level of TROG and BPVS VMA.

Happé's (1995) study with 70 participants with ASD found that there was a BPVS VMA lower threshold of 5.5 years, and an upper threshold of 11.58 as regards passing FB tasks. The findings from this study are about 6-8 months lower in both cases (although that lower threshold is reliant on only one child).

Happé found no correlation between BPVS performance and FB in children with MLD. The present study did not replicate this result, finding highly significant correlations between BPVS and FB. This difference is not simply due to a larger sample size, as the magnitude of the correlation is greater, not just the significance level. However, it is true that no clear relationship between either vocabulary or grammar and FB performance was found in the MLD group – although there were

significant positive correlations between the language measures and FB, logistic regression models did not predict more than 65% of cases correctly.

The findings from this study support claims by Happé (1995) and Tager-Flusberg and Sullivan (1994) that language may be particularly important in the development of FB understanding in individuals with ASD. The strong associations between the language means and FB within the ASD group cannot be due to linguistic demands of the FB tasks – TD children pass these tests by the age of 4, yet a child with ASD appears to need a level of grammar nearly 2 years above that to have even a chance of passing. The existence of a third factor influencing both language and FB is a possibility, although the most plausible candidates – such as joint attention – are so closely linked to ToM that to separate their influences seems impossible. It appears that the relationship between language and ToM is likely to be a genuine one, and it is interesting to consider what might be its cause. A relationship between vocabulary and FB understanding tells us little about the causal direction of the relationship. It is possible that a good vocabulary allows children to develop an understanding of FB, or alternatively it is possible that a child cannot achieve a good vocabulary without some understanding of mind or mentalising ability (Bloom, 2000; Happé, 1995). A deficit in ToM and its precursors (such as joint attention) could prevent vocabulary gains. However, although it may be necessary to have a certain level of grammatical knowledge to acquire an understanding of FB, it is less easy to see how a deficit in ToM would prevent grammatical gains except by delaying general language acquisition, in which case vocabulary and grammar should be equally affected. The special association between grammar and FB performance found here, and elsewhere (Tager-Flusberg & Sullivan, 1994), therefore indicates that some elements of language may be necessary for FB success in children with ASD, rather than vice versa. The two options are not exclusive however, as early forms of social understanding may be necessary for early language gains, which are then themselves necessary for ToM development.

This begs the question of how language might be influencing ToM development in ASD. Children who pass ToM tasks still demonstrate the characteristic social impairments associated with ASD, their ability to pass FB tasks does not mean that their social difficulties are solved, although there is evidence that some of those who pass FB tasks show more socially insightful behaviour such as taking hints, lying and cheating (Frith et al., 1994). It therefore seems likely that

some, at least, of these children have not simply learnt to pass FB tasks, but do actually have a better understanding of minds than do most children with ASD, albeit not a good enough understanding to completely overcome their social and communication difficulties. Their performance will be investigated further in studies 4 and 5 (chapters 8 and 9).

How might language influence ToM development in ASD? It is possible that language skills give children with ASD access to social experiences and interactions that enable them to develop a basic ToM. However, this leaves unanswered the question of why children with ASD should need better language skills than TD children to access these experiences. Alternatively, language itself may provide a structure for developing representational understanding. Happé (1995) and Tager-Flusberg (2000) suggest that children with ASD may use cognitive strategies to solve FB tasks in a way that is not necessary for TD children (or presumably children with MLD), and that language may provide the structure for this logical reasoning.

As the TROG does not contain any mental state or communication verbs our data does not directly test Tager-Flusberg's theory that the complement structures of communication verbs are particularly important for children with ASD. However, we did find that children with ASD required levels of language up to two years higher than TD children to have a chance of passing FB tasks, and that language appears to be more strongly related to ToM in ASD than in MLD. The role of alternative strategies in passing ToM tasks in ASD will be further examined in chapter 8.

## **Summary**

This study found that children with ASD were more likely than those with MLD to fail FB tasks. However, a significant proportion of children with MLD also failed these tasks. The performance of those with MLD appeared to be more affected by task demands, and there was evidence that they found it easier to remember their own FB than to predict another person's. Language, especially grammar, was strongly related to FB performance in ASD. These findings provide support for the theory that language is necessary for ToM development, particularly in ASD, where it may provide an alternative neuro-cognitive route to representational thought. In those with MLD, there was evidence that they may be failing FB tasks for reasons other than a lack of a ToM. Possible sources of difficulty are executive problems, and these will be discussed in detail in the next chapter.



## **CHAPTER 4**

### **THE RELATIONSHIP BETWEEN EXECUTIVE FUNCTION AND THEORY OF MIND**

Chapters 2 and 3 have discussed the relationship between language and ToM in children with ASD and MLD. Another factor that has been linked to ToM performance is executive functioning (EF). This chapter will review the evidence for a relationship between ToM and EF in normal development and in those with developmental disorders, and chapters 5 and 7 will investigate the relationship in children with ASD and MLD experimentally.

Much recent work has focused on the relationship observed between performance on tasks of EF and ToM. Several studies have demonstrated links between these tasks, both in young TD children, and in children with developmental disorders. The concept of a 'theory of mind' has been discussed in depth in chapter 1. The processes included under the umbrella term of 'executive function' will briefly be described here, before discussing possible relationships between these two competencies.

Executive function, as defined for example by Perner (2000), is a term for processes responsible for higher level action control, in particular those that are necessary for maintaining a mentally specified goal and for implementing that goal in the face of distracting alternatives. Executive functions enable an individual to carry out novel actions and to act on their knowledge and understanding of a situation. Deficits in executive functioning may therefore lead to widespread difficulties, both in real life and on experimental tasks, and may be responsible for poor performance on tasks where participants understand the concepts involved, but do not have the necessary level of self-control to demonstrate their competence.

Processes subsumed under the term executive function include inhibition, set shifting, planning, coordination and control of action sequences. Tasks tapping set shifting include the Wisconsin Card Sort Task (WCST:Heaton, 1981), and the Dimensional Change Card Sort (DCCS: Frye et al., 1995), both of which require the participant to sort cards by a variety of rules. Examples of planning tasks are the Tower of Hanoi and Tower of London, in which the child must copy an arrangement of disks or balls in a restricted number of moves, with certain rules restricting their

movements. Inhibition tasks include the Go-NoGo task (Drewe, 1975), in which children must respond to a certain stimulus (e.g. a plane) whilst inhibiting their response to another (e.g. a bomb). Tasks of working memory include the Backwards Digit Span, where the participant must repeat a lengthening series of digits backwards. Many EF tasks (and particularly those which relate to performance on FB tasks) require suppression of a prepotent or habitual response in favour of a new response (e.g. WCST, Go-NoGo), and failure of EF in these tasks is therefore manifested as perseveration of the old response.

A growing number of studies have demonstrated a relationship between EF and ToM both in young TD children, and in those with developmental disorders. Several theoretical explanations for this relationship have been suggested. In this chapter the evidence for a relationship between EF and ToM will be summarised, both in TD children and in those with developmental disorders, and the theories will be discussed in light of the available evidence.

### **EF and ToM in young typically developing children.**

It is well established that children pass tests of FB understanding (and thus are often credited with the possession of a ‘theory of mind’) at around the age of 4 years (see chapter 1). Their understanding of mind appears to start to develop much earlier, however. By the age of eighteen months children appear to have an understanding of goal-directed action and some understanding of intention (Meltzoff, 1995) and understanding that different people may have different preferences (Repacholi & Gopnik, 1997). From the age of 2 years children begin to demonstrate their understanding of perspective taking through speech. Bartsch and Wellman (1995) reported older 2-years-olds explicitly contrasting different people’s thoughts and feelings. However, the explicit understanding of mistaken beliefs (and therefore the ability to lie and deceive) does not appear to emerge until the age of around 4 (Wellman, Cross et al., 2001). At around the age of 4, children begin to show a clear understanding of mistaken belief as shown in a number of different tasks.

At around the same age (4 years) as the majority of children pass FB tasks, young children improve substantially on EF tasks requiring the inhibition of a prepotent response (e.g. an habitual rule) in favour of a new response (Russell, Mauthner, Sharpe, & Tidswell, 1991). An example of the type of task which 3-year-olds find difficult is a simplified version (Hughes, 1998a) of Luria’s hand game

(Luria, Pribram, & Homskaya, 1964), in which children are first asked to copy the experimenter's hand actions (e.g. a fist or a pointed finger), and then to do the opposite action to the experimenter. They thus have to inhibit the rehearsed tendency to copy the action. Three-year-olds also perform badly on a range of other tasks which share this need to inhibit a prepotent tendency, including the DCCS (Frye et al., 1995) and a highly simplified version of the WCST (Hughes, 1998a). In both of these tasks, the child must suppress a tendency to sort the cards by a previous rule, and instead sort by a new rule. Another task which improves at around the age of 4 years is the backwards digit span (Davis & Pratt, 1995), in which children must repeat a list of numbers in reverse order, thus inhibiting the natural tendency to repeat the numbers as heard.

This improvement in executive functioning at the age of around four has been shown to be correlated with ToM development, even when chronological and mental age are partialled out (e.g. Carlson & Moses, 2001; Hughes, 1998a). Table 4.1 summarises those studies with young normal children using tests of EF and ToM. For ease of comparison the correlations given are mostly averages of the more specific relationships between tasks cited in the studies. Separate correlations are given when they are particularly informative.

**Table 4.1. Mean correlations from studies looking at the relationship between ToM and EF tasks in TD children**

Authors	Age of children	ToM tasks	EF tasks	ToMxEF (age partialled out)
Russell (1991)	3:0-4:10 (n=33)	FB UT prediction	Windows task	0.89 (0.70)
Davis & Pratt (1995)	3:3-5:4 (n=54)	FB DB self and other, false photo task.	Backwards digit span	0.46
Frye (1995) Experiment 1	3:1 – 5:5 (n=60)	A-R, FB DB self and other	DCCS card sort, physical causality task.	0.38 (0.16)
Experiment 2	2:8-5:4 (n=40)	A-R, FB DB self and other	DCCS card sort	0.50 (0.37)
Experiment 3	2:9-6:1 (n=60)	A-R, FB DB self and other	Card sorting task in 3 version	0.33 (0.25)
Gordon & Olsen (1998)	3:0-6:4 (n=72)	A-R, FB DB self and other	Counting and labelling task, finger tapping and labelling task.	0.56 (0.46)
Hughes (1998a)	3:3-4:7 (n=45)	FB UT prediction, FB UT explanation, deception	6 EF test battery (e.g. hand game and card sorting)	0.49 (0.30)
Hughes (1998b)	Time 1 3:3-4:7 (n=45)	FB UT prediction, FB UT explanation	5 EF test battery (e.g. hand game and card sort)	0.27
	Time 2 4:2-5:8	2 <sup>nd</sup> order FB	As above	0.34

Authors	Age of children	ToM tasks	EF tasks	ToMxEF (age partialled out)
Hughes, Dunn & White (1998)	3:6-4:6 (n=40)	10 FB tasks, including UT, DB and deception	6 EF test battery including IC, WM, planning and attention.	0.31 (-.19) <sup>a</sup>
Roth & Leslie (1998)	3:0-5:0 (n=57)	FB UT prediction	IC: Screen task	0.38
Perner (1998)	3:0-5:11 (n=57)	FB UT prediction	DCCS card sort	0.59 (0.48)
Cole(2000) Experiment 1	3:2-5:1 (n=121)	FB DB self and other, A-R	4 EF test battery (including DCCS)	0.38(DCCS only)
Cole(2000) Experiment 2	3:11-5:8 (n=71)	FB DB self and other. FB UT prediction,	DCCS, inhibitory control	0.33(DCCS only)
Carlson (2001)	3:3-4:11 (n=107)	A-R, FB DB self and other, FB UT prediction, deceptive pointing,	10 EF test battery (including DCCS)	0.66 (0.34)
Charman, Carroll & Sturge (2001)	8:1 – 10:10 (n=22)	ToM stories	ToH (planning) and Go-NoGo (IC)	Planning: 0.29 (0.32) <sup>a</sup> IC: 0.43 (0.38) <sup>a</sup>
Carlson, Moses & Breton (2002)	3:4-5:6 (n=47)	A-R, FB DB self and other, FB UT prediction	6 EF test battery testing IC and WM	IC:0.41 (0.18) <sup>a</sup> WM:0.38 (0.03) <sup>a</sup>
Perner, Lang & Kloo (2002) Experiment 1	3:1 – 6:2 (n=56)	FB UT explanation, FB UT prediction	DCCS	FB Prediction 0.65 (0.40) <sup>a</sup> FB Explanation 0.61 (0.38) <sup>a</sup>
Perner, Lang & Kloo (2002) Experiment 2	2:9-5:8 (n=73)	FB UT explanation FB UT prediction	DCCS IC: Go-NoGo	DCCS: 0.50 (0.28) <sup>a</sup> IC:0.24 (.00) <sup>a</sup>
Perner, Kain & Barchfeld (2002)	4:9-6:8 (n=22)	2 <sup>nd</sup> order FB	11 EF battery including attention, planning, WM and IC tasks.	0.39 (.31) <sup>a</sup>

<sup>a</sup>correlation with age and IQ partialled out Abbreviations: AR = Appearance-Reality task. FB = False belief, UT = unexpected transfer task, DB = Deceptive box task, DCCS = Dimensional-change card sorting. IC = inhibitory control. WM = working memory. ToM stories = Happé's Strange Stories (Happé, 1994). ToH = Tower of Hanoi

Perner and Lang's (1999) useful meta-analysis of nine recent studies with TD children aged between 2:8 and 6:4 found a strong overall effect size for the relationship between EF and ToM. The general pattern seen from this table is one of a strong positive correlation between performance on FB tasks and on EF tasks, even when the EF tasks are treated as a single entity. This is particularly clear with tasks that require the inhibition of a prepotent response – both in tests of attentional flexibility such as the DCCS, and tests of inhibitory control such as the Windows task and Luria's Hand Game. Carlson et al. (2002) found a specific relationship between FB and inhibitory control requiring the child to chose between two conflicting responses, as opposed to the inhibition involved in delaying gratification. There are also a number of studies showing a relationship between working memory and

performance on FB tasks. A possible explanation for this may be because of working memory demands of keeping one perspective in mind whilst considering another perspective.

### **EF and ToM in children with developmental disorders.**

As discussed in chapter 1, there is extensive evidence indicating the existence of a ToM impairment in children with ASD. There is also evidence for deficits in EF in this group (see Sergeant et al., 2002 for a recent review). However, whether these deficits are specific to ASD is not clear. Executive function deficits are seen in a range of developmental disorders, for example Attention Deficit Hyperactivity Disorder (ADHD) (Grodzinsky & Barkley, 1999), Tourette's Syndrome (Harris, Schuerholz, Singer, Reader, & et al., 1995) and conduct disorder (Morgan & Lilienfeld, 2000). It is not yet clear what distinguishes the EF problems in ASD from those seen in other groups. Hughes (2001) suggests that they may characteristically involve high-level and non-spatial problems of inhibition, whilst Ozonoff (1999) points to difficulties in flexibility and set shifting. Table 4.2 summarises the findings of studies of EF and ToM in individuals with developmental disorders.

***Table 4.2. Studies looking at the relationship between ToM and EF tasks in those with developmental disorders.***

<b>Authors</b>	<b>Participants</b>	<b>ToM tasks</b>	<b>EF tasks</b>	<b>EF x ToM</b>
Ozonoff (1991)	23 ASD CA: 8:1 – 20:11 20 controls with dyslexia, ADHD, MLD. CA = 8:7-19:6	ToM battery of 5 items, including A- R . FB DB (other), 2nd order FB	Tower of Hanoi, WCST	ASD 1 <sup>st</sup> order FB=0.64 2 <sup>nd</sup> order FB=0.50 Non-autistic 1 <sup>st</sup> order FB=0.13 2 <sup>nd</sup> order FB=0.39 .
Baron-Cohen and Robertson (1995)	1 ASD CA = 13:6, 1 GTS CA=13:1 1 co-morbid ASD and GTS CA = 13:5	FB UT and DB (other), Deception	Intention editing: Yes and No game, Luria Hand Game, Day-Night Stroop Fluency: FAS test	ASD: poor ToM, but good intention editing. GTS:good ToM, poor intention editing. ASD and GTS: poor ToM and intention editing. All poor on fluency test.
Zelazo (1996)	12 Down Syndrome CA =16:0 – 30:9 12 TD preschoolers CA= 5:2 – 6:8	FB DB (other and self), A-R, pretend reality.	DCCS	Correlations with both groups combined $r = 0.59 (0.51)^a$
Tager-Flusberg (1997)	10 Praeder-Willi syndrome, CA = 5:6 – 8:8 14 William's Syndrome CA = 5:5 – 8:11	FB UT prediction	Flexibility: Day- night Stroop, Tapping task	Non-significant relationship between passing and failing across both groups.

Authors	Participants	ToM tasks	EF tasks	EF x ToM
Hughes, Dunn and White (1998)	40 'Hard-to-manage' preschoolers CA = 3:6-4:6	10 FB tasks, including UT, DB and deception	6 EF test battery including IC, WM, planning and attention.	0.32 (0.35) <sup>a</sup>
Roth and Leslie (1998)	21 ASD CA = 6:0 – 19:1	FB UT prediction	IC: Screen task	No relationship (r = 0)
Garner (1999)	8 Fragile-X CA = 10:3 – 14:2 8 MLD CA = 12:4 – 14:9	FB DB self and other, FB UT prediction, 2 <sup>nd</sup> order FB task.	Modified WCST	Not reported, deficits in both groups.
Charman, Carroll and Sturge (2001)	22 Boys with ADHD CA = 6:7-10:6	ToM stories	Tower of Hanoi (planning) and Go-NoGo (IC)	Planning: 0.07 (0.06) <sup>b</sup> IC: .06 (.18) <sup>b</sup>
Perner, Kain and Barchfeld (2002)	24 preschoolers at risk of ADHD CA = 4:7-6:6	2 <sup>nd</sup> order FB	EF battery of tasks including IC, planning and attentional tasks	0.14 (0.16) <sup>a</sup>
Zelazo, Burack and Frye (2002)	10 medium functioning CA = 9:3 – 27:1 12 lower functioning CA = 7:7- 16:2 individuals with ASD	FB UT prediction Explicit FB	EF battery of tasks requiring rule-use including DCCS	Lower-functioning r = -0.40 Medium-functioning group r = 0.82

<sup>a</sup> with ability partialled out <sup>b</sup> with age and ability partialled out

Abbreviations used: FB = False belief, A-R = appearance-reality, DB = deceptive box, UT = unexpected transfer, WCST = Wisconsin Card Sort Task, DCCS = Dimensional Change Card Sort task, IC = inhibitory control, ToM stories = Happé's Strange Stories (Happé, 1994), GTS = Gilles de la Tourette Syndrome.

The findings here are less consistent than those with young TD children – perhaps unsurprisingly, given the variety of disorders and ages involved in the various studies. Several studies have reported a strong association between EF and ToM task performance in higher functioning people with ASD. A single study (Roth & Leslie, 1998) found no association between an executive task and FB in ASD, this study will be discussed below. The findings from other developmentally disordered groups are less consistent. The Zelazo (1996) study, which indicated a possible connection in Down Syndrome (DS) reports only the correlation of both the DS group and young preschoolers combined. In ADHD or those at risk of ADHD there appears to be a relationship in preschoolers that becomes less significant with age. A less comprehensive range of tasks has been used with those with developmental disorders than with young TD samples, and as such it is hard to distinguish what aspects of EF might be related to FB in these groups. However, all the studies that report high correlations used some form of card sorting task in their EF battery, and the results indicate that set shifting and flexibility may be particularly related to ToM. Those

using inhibitory control tasks (Charman, Carroll et al., 2001; Perner, Kain et al., 2002; Roth & Leslie, 1998) found much lower correlations than those seen in young TD children.

### **Why is there a relationship between EF and ToM?**

Several explanations have been suggested for this relationship. Five major alternatives are outlined below. This chapter will only discuss the first four since the fifth theory is on a neurological rather than a psychological level and as such is not a focus for this thesis. It is also possible for it to co-occur with some of the other theories. The five alternatives are:

- 1. ToM tasks have an executive component that causes children to fail.**
- 2. ToM is a prerequisite for (or an integral part of?) the development of EF.**
- 3. EF is a prerequisite for ToM.**
- 4. A third factor underlies both EF and ToM (for example, a common logical structure).**
- 5. Common brain structures mediate both EF and ToM.**

### **ToM tasks have an executive component that causes children to fail.**

The standard unexpected transfer FB task has an obvious inhibitory component. To pass the test, children must suppress their knowledge of reality, and predict that the protagonist will look in the empty box, usually by pointing. This led to suggestions that it might be this inhibitory requirement that causes children to fail, rather than the ToM demands of the tasks (Russell et al., 1991). Russell (Hughes & Russell, 1993; Russell et al., 1994; Russell et al., 1991) used a 'Windows task' to assess children's ability to inhibit a prepotent response. In this task the child sees two boxes with windows in them. One box contains a sweet, the other is empty. To win the sweet, the child must point to the empty box. They are trained on the contingences beforehand. However, despite this training, most 3-year-olds and children with ASD pointed to the full container, and continued to do so for 20 test trials, despite never being rewarded for this strategy. Initial forms of the task had an opponent, who won the sweet when the child pointed to the full box (and thus the task could have been said to have a ToM/deceptive element). However, later forms of the task carried out without an opponent proved just as difficult for 3-year-olds and children with ASD. This appeared to be strong evidence for inhibitory difficulties that could account for

the poor performance of most 3-year-olds and children with ASD on the standard FB task.

Leslie and colleagues (German & Leslie, 2001; Roth & Leslie, 1998; Scholl & Leslie, 2001) put forward a different account of the role of EF in FB task performance. They suggest that, in order to pass FB tasks, it is necessary to have both a functioning ToM mechanism (ToMM) and a selection processor (SP) (see chapter 1 for an outline of Leslie's theory of ToM). The ToMM is defined as a module that spontaneously processes behaviours that are attended to, and computes the mental states that contribute to those behaviours, whilst the selection processor is a general executive function required in many situations to inhibit salient but unwanted responses (Scholl & Leslie, 2001). Roth and Leslie (1998) suggest that 3-year-olds possess the concept of belief, but have difficulty in predicting behaviour on the basis of that belief. Their first study included a 'partial true belief' task, which paralleled the unexpected transfer FB task, except that instead of the original object being moved, another identical object was produced and placed in a different location. The children were asked whether the protagonist knew that there was a coin in the new location, and then asked where the protagonist would look for the coin on their return. Seventy percent of the 3-year-olds in their study passed the 'knowledge' question, but only 33% of these then predicted behaviour correctly on the basis of that knowledge. This contrasts with Leslie and Frith's (1988) findings on the same task with children with ASD, in which 61% of the children passed the knowledge question, and 73% of these then went on to correctly predict behaviour. Roth and Leslie point to these results as evidence that 3-year-olds and children with ASD fail standard FB tasks for different reasons.

Roth and Leslie's (1998) 'Screen' task, designed to put demands on the selection processor without taxing the ToMM showed no correlation with performance on FB tasks in children with ASD, but was highly significantly correlated in three-year-old children. This task involves an opaque screen, a basket, a box and a marble. The marble is placed in the basket, and the display is placed behind the screen, out of the children's view. Another identical set of basket, box and marble is then placed in front of the screen, and the marble is moved from the basket into the box. The child is then asked where the marble is behind the screen, and where it is in front of the screen. Three-year-olds found the Screen task nearly as difficult as standard FB, and more difficult than non-standard FB tasks. Children with



ASD, in contrast, were almost at ceiling on the screen task, whilst more than 50% of them failed the non-standard and 65% the standard FB task. Roth and Leslie interpret this as evidence that children with ASD have an intact SP, but impaired ToMM, whilst the 3-year-olds had an immature SP, but intact ToMM. However, the correlation between FB and performance on the Screen task in 3 to 4-year-olds is not particularly high ( $r = .38$ ). If limitations of the selection processor account for *all* the difficulties demonstrated by young children on FB tasks, this correlation might have been expected to be higher.

Leslie's account differs from Russell and colleagues' in that he contrasts TD 3-year-olds (who he claims fail FB tasks due to an immature selection processor, but have an intact ToMM), with children with ASD, who he argues fail due to an impaired ToMM, despite having an intact SP. This would seem to imply that the relationship between performance on FB tasks and inhibitory control and flexibility should only be evident in TD children. It would also follow that if we could reduce the executive demands of FB tasks, not only would TD children pass at an earlier age, but also that the association with EF should be reduced in normal samples.

It does seem to be the case that reducing the executive demands makes FB tasks easier for young TD children. Adding a photographic cue (whereby the participant posted a photo of what they thought was in the box before it was opened) aided performance on the Deceptive box task in TD children and children with ASD (Charman & Lynggaard, 1998), and telling the young TD children of the new location of an object rather than showing them improved performance on the unexpected transfer task (Zaitchik, 1991). However, in spite of these manipulations, the Wellman et al. (2001) meta-analysis found that children aged below 3 years 6 months performed at best at chance level on FB, whilst children older than 4-years-old passed, indicating that executive demands cannot account for the developmental transition seen between the ages of 3 and 4 years.

One type of task which in fact was not included in the Wellman et al (2001) meta-analysis, but which appears to substantially reduce executive demands of the task is the 'FB explanation' task in which the protagonist returns and searches in the wrong box, and the child is asked to explain why they searched there. The demands on the SP in a task like this are presumably minimal, as the selection has effectively already been made. However, children pass FB explanation tasks at around the same time as they pass FB prediction tasks, and the explanation tasks correlate as highly

with EF tasks as do the prediction tasks (Hughes, 1998, see table 4.1 for summary), and show the same developmental trend. These findings appear problematic for the ToMM/SP account, as there are no apparent demands on the SP, and yet 3-year-olds still fail, indicating that their difficulties in FB tasks go deeper than difficulty with inhibiting a salient response. In addition, they suggest that the relationship between EF and ToM is deeper than one due simply to task demands.

There is less controversy over the case of children with ASD. Not only do they pass tasks that aim to control for the inhibitory demands of the FB task (the False Photo and Screen task), but there is myriad evidence from tasks with no obvious executive demands that they do have ToM deficits (see chapter 1 of this thesis, Baron-Cohen, 2000; Happé, 1994). Russell and colleagues, who initially proposed that executive factors were responsible for FB failure in children with ASD, now favour a more developmental account, in which early EF is a precursor for ToM (Russell, 1997b).

However, despite it seeming improbable that task specific executive factors are entirely responsible for FB task failure in either young TD children, or children with ASD, they may be one source of individual differences in performance (Wellman, Cross et al., 2001). It is possible that in a cohort of young 3-year-old children, those with superior inhibitory control will have an advantage in FB tasks, and all the children will show an improved performance on a task with fewer inhibitory demands. These improvements will be due to the lack of the systematic errors typically seen in young children, which will mean their performance improves from below-chance to chance level. At a more transitional stage (in terms of understanding of belief), those children with poor inhibitory control may perform at chance, whilst those with better control perform above chance. Here, Leslie's analysis is useful, in that it identifies two distinct components to passing a FB task, both of which may be of varying importance at different times in development. However, this analysis ignores the question of whether these two components are related, and if so, how.

To briefly summarise, it seems unlikely that executive demands of the FB task mask the ability to attribute mistaken thoughts or beliefs either in young normally-developing children, or in children with ASD. However, executive factors may account for individual differences in performance in young TD children in the year

before they pass FB tasks. It seems very likely that the relationship between EF and ToM is not simply due to task demands.

**ToM is a prerequisite for (or an integral part of?) the development of EF, or EF is a prerequisite for ToM**

The ToM as a prerequisite for EF theory has two forms. An early and strong form suggests that it is necessary to have an understanding of metarepresentation and mentality in order to gain control of mental processes and actions, and that therefore without some understanding of mental states, the development of EF is not possible (Wimmer, 1989). A later and somewhat weaker form (Perner et al., 1999) suggests that ToM is an integral part of EF. Perner's argument goes like this: In order to inhibit unwanted actions, the child must understand that there are existing schemata which make one act in an unwanted way unless one actively inhibits those actions. Understanding the existence of these schemata as entities which make one act in a certain way is to understand them as representations with causal powers, and this understanding can therefore be dubbed metarepresentational. However, it is not clear from Perner's account how much of this metarepresentational understanding is conscious. According to this theory, an understanding of the causal/representational nature of mental states is necessary for a child to develop *executive* inhibition, which they define as inhibitory control directed at action schemata on the basis of their representational content. Perner contrasts this with general inhibition, which he suggests can happen at a lower level of contention scheduling. This will be discussed in greater detail below, alongside evidence for the third option.

Russell (1996; 1997b) suggests that very early executive control, in the form of action monitoring and agency, is a prerequisite for the development of self awareness and therefore the later development of a ToM. This theory has the strong implication that early difficulties in EF should by necessity lead to later difficulties in ToM.

The evidence for these two theories (the functional dependency theories) will be discussed together, as many of the studies which provide evidence for one side are therefore evidence against the other, and many points are relevant to both.

Many children with ASD who show ToM deficits also show severe impairments in EF. Pennington and Ozonoff (1996) reviewed 14 studies that measured EF in children with ASD. Thirty-two different tasks were used, and

children with ASD were impaired (in contrast to ability matched controls) on 25 of them. Sergeant et al. (2002) reviewed a range of studies which showed a deficit in EF in children with ASD when compared to TD controls. However, it is less clear whether there is a deficit when compared to children with other developmental disorders such as ADHD or MLD. As mentioned above, some researchers have suggested that it is in fact essential to be able to reflect on mental states in order to develop executive function (e.g. Wimmer, 1989). In its strongest form, this theory suggests that it should not be possible to have a ToM deficit whilst demonstrating no EF deficit, and that early difficulties in ToM should reliably predict later difficulties in EF. A later form of this theory (Perner et al., 1999) argues that a ToM is an integral part of EF. Hughes (1998) found that in 50 TD children, early performance on EF tasks (at age 3:7) accounted for 34% of the variance in ToM tasks at age 5. When age, verbal ability and initial ToM were taken into account, performance on EF task still predicted 20% of the variance in ToM at age 5. In contrast, early ToM measures correlated significantly with only one later EF task (although all the correlations were positive) and a stepwise regression showed no significant predictive relationship between early ToM and this task. This appears to support an account by which EF is necessary for later ToM development, rather than vice versa.

Ozonoff et al (1991) assessed EF and ToM in a group of 23 high-functioning children with ASD, and found that impairment on the EF tasks (when compared to an IQ-matched group with learning difficulties) was greater than on the ToM measures, and that the EF impairment was a better discriminator between the groups. The authors conclude that the EF deficit is more likely to be the primary cause of ASD, as it discriminates better between the groups. However, there are a number of problems with this. The finding may well have been due to the choice of ToM tasks (1<sup>st</sup> and 2<sup>nd</sup> order FB tasks); both groups were almost at ceiling on the 1<sup>st</sup> order FB tasks, which restricted the potential for any group difference. When the groups are compared on the 2<sup>nd</sup> order task alone, then ToM performance discriminated between the groups as well as did the EF measures (Perner, 2000). It may be that the FB tasks used were not appropriate for the higher functioning members of the ASD group (whose IQs ranged from 55 to 140), and a more complex mentalising task would have revealed more pervasive difficulties. From a developmental perspective, if it could be shown that only some children with ASD have ToM difficulties, whilst all have EF difficulties, this would suggest that deficits in ToM are not the necessary cause of EF problems.

However, Ozonoff et al. did not take a developmental perspective, and since the ToM tasks in this study may not have been demanding enough for the very high functioning members of the group, we cannot conclude this from their findings. It is also not clear whether the ToM and EF tasks used were comparably difficult.

A study by Baron-Cohen and Robertson (1995) documents 3 single cases of 13-year-old children, one individual with autism, one with Gilles de la Tourette's Syndrome (GTS) and one with both. The children were given 3 ToM tasks (1<sup>st</sup> order FB, unexpected transfer and deceptive box, and the penny-hiding deception task), 3 'intention editing' tasks (executive tasks mostly requiring inhibitory control and flexibility; for example the day/night Stroop task), and a further test of generativity, the FAS verbal fluency task. The child with autism failed all the ToM tasks whilst succeeding on all of the 'intention editing' EF tasks, whilst the child with GTS failed 2/3 EF tasks but succeeded on all ToM tasks. The child with both GTS and autism performed poorly on both EF and ToM measures. All three children performed poorly on the verbal fluency test. This would appear to indicate a dissociation between EF and ToM, with neither being essential for the development of the other. However, there are several problems with this approach. Firstly, there are obvious difficulties with extrapolating too much from single cases. Secondly, this assumes that both ToM and the inhibitory component of EF are unitary concepts that a child either has or does not have. Whilst it is certainly the case that a child can pass or fail a ToM task, this may be a limitation of the tasks rather than a reflection of the nature of ToM as a unitary concept which a child either possess or doesn't possess. In the case of EF tasks, there is no natural pass or fail mark, and so these are set somewhat arbitrarily. It is entirely possible that, whilst EF is important for the development of ToM, a lower level of inhibitory control is required for the development of ToM than has been set as the pass mark in most studies. It is also the case that the tests that were used in the Baron-Cohen and Robertson study (1<sup>st</sup> order FB tasks and a deception task) are likely to have been too easy for the participant with GTS (who had a VMA of 9 years). Using tasks more appropriate for that age group might have revealed deficits.

Tager-Flusberg (1997) tested 10 children with Prader-Willi syndrome and 14 children with Williams Syndrome on two FB and two EF tasks. Perner (2000) gives some further details of their data that indicate that, whilst 6 children (3 from each group) passed both ToM tasks whilst failing both EF, no children passed both EF

tasks whilst failing both ToM. Looking at those children who did better in one area than another (but who did not fail or pass both in each area) reveals four cases where the children did better on ToM than on executive functioning, but three cases where children did better on EF than on ToM. The strongest evidence (the dissociations seen in 6 children) would therefore appear to indicate that intact EF is not necessary for intact ToM.

A further difficulty for the theory that EF is fundamental to the development of a ToM is the existence of clinical disorders such as ADHD, obsessive compulsive disorder (OCD) and Tourette's Syndrome (TS), in which individuals demonstrate executive dysfunction without showing autistic-like difficulties in ToM. Perner et al. (2002) in a longitudinal study of children with behavioural problems, found pervasive problems in executive functioning at age 4.5 – 6.5, but no deficit in second order ToM tasks at age 6. This is problematic for an account that claims that intact EF is necessary for ToM development.

To summarise, there are studies that could be used to support the theory that a ToM is necessary for or an integral part of EF development (e.g. Perner, Kain et al., 2002; Tager-Flusberg et al., 1997), and there are also studies which could be used to argue that EF is necessary for ToM development (e.g. Hughes, 1998a). In addition, there is some evidence for a dissociation between the two (Baron-Cohen & Robertson, 1995). In order to further elucidate the nature of the relationship, longitudinal studies are needed. However, given the logistical problems with this type of study, an alternative would be intervention studies which trained children in either ToM or EF, and which tested them in both domains before and after training, in order to identify any interactions. Children with ASD would be particularly appropriate for this type of training, since they demonstrate difficulties in both domains, and because it is particularly difficult to do longitudinal studies with young children with ASD, as they are typically not diagnosed until the age of 3 or 4 years at the earliest, and are often late in developing language and other skills which would be necessary in order to test them in the usual way. Chapter 7 will report the findings of such a training study.

### **EF and ToM tasks share a common logical structure**

Zelazo, Frye and colleagues (Baron-Cohen & Robertson, 1995; Frye et al., 1998) suggest that the relationship between tasks such as the DCCS and ToM is seen

because the development of both depend on underlying domain-general mechanisms, specifically rule use involving an embedded conditional ‘if-if-then’. They use the DCCS task, in which children must sort cards that vary along two dimensions (for example colour and shape). Initially participants must sort by one dimension, and then after a switch (which are told about) they must sort by the second dimension. The authors argue that the DCCS involves this type of rule as follows: *if* we are playing the colour game, and *if* you give me a green circle, *then* I will put it with the green square target. However, *if* we are playing the shape game and *if* you give me a green circle, *then* I will put it with the yellow circle target. In the unexpected transfer FB task, the rule could be analysed as follows: *if* Sally did not see us move the marble, and *if* she wants to find the marble, *then* she will look in the wrong box. They argue that the ability to pass FB tasks reflects the use of these rule systems, and that this accounts for the association seen between FB performance and performance on the DCCS (and, presumably, other EF tasks).

Zelazo, Frye and colleagues (e.g. Frye et al., 1998; Zelazo, Carter, Reznick, & Frye, 1997) suggest that the correlation between EF and ToM in normal development, and the co-occurrence of such difficulties in ASD, are accounted for by domain-general demands of reasoning, specifically embedded conditionals or the ability to integrate two opposing perspectives. They argue that the developmental transition seen at age 4 is due to the child's emerging ability to reason with these embedded conditionals. Zelazo et al (2001) suggest a domain-general developmental framework for executive functioning with three key dimensions of cognitive development: consciousness, rule complexity and behavioural control. They argue that developmental changes in reflexive consciousness permit the formulation and use of increasingly complex rule structures, which in turn permit increased flexibility and control. In these terms, they argue that the deficits in ToM seen in people with ASD can be understood as a problem with using rule systems, and are therefore fundamentally difficulties in executive functioning. However, it appears possible that a ‘reflexive consciousness’ is in itself part of what others would consider a ToM – this is not well defined in Zelazo et al.’s account.

### **Evidence for and against**

Zelazo (1996) found a strong association between performance on FB tasks and the DCCS in individuals with Down Syndrome (DS), with the individuals with DS focusing on a single state of affairs in both tasks (reality in the case of the FB

tasks, the first rule in the case of the DCCS). Frye (1995) found a similar relationship in TD 3 to 5 year olds, and Zelazo et al. (2002) found a high correlation between performance on the DCCS and FB tasks in high-functioning individuals with ASD. In addition, Zelazo et al (2001) point to the growing body of evidence that a ToM deficit (at least when measured by FB tasks) is neither unique to ASD nor specific to the ToM domain (see Yirmiya et al. 1998 for a meta analysis), as evidence for a more general cognitive deficit. Others have not always found the same deficits, however, Baron-Cohen et al. (1985), for example, found good performance on FB tasks in their control group, who had Down Syndrome.

However, Perner (1999) argues that the embedded conditional structure which Zelazo and colleagues postulate underlies difficulties in FB tasks can just as well be applied to a food-preference task (Repacholi & Gopnik, 1997) which children can pass as young as 18 months. This task could fit the if-if-then format as follows. *If* you want food, and *if* you like broccoli, *then* I will give you broccoli (as opposed to biscuits, the child's own preference). Perner also argues that the application of this rule structure is arbitrary, and FB tasks can be just as easily cast in an *if-then* structure; e.g. *if* Sally is looking for her apple *then* cupboard. In addition, whilst embedded conditionals may encapsulate a similarity between FB and the DCCS task, it is not clear how it can account for the widespread social difficulties seen in people with ASD. In more advanced tests of mentalising such as the Eyes task (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997), the participant is asked to identify mental states from a picture of eyes. This task requires no embedded conditionals, and thus if this was the key to the problems demonstrated by people with ASD, we should expect no difficulties. However, children with Asperger's Syndrome found this task substantially more difficult than did controls with Tourette's Syndrome. Zelazo et al (2001) admit that, in addition to a grasp of the embedded conditional, 'mental state concepts will perhaps be useful' (p.206) when reasoning about human behaviour. This theory is silent on the question of how these mental state concepts emerge, and whether they might also be impaired in people with ASD. There are cases of children performing well in ToM whilst failing tests of EF (Baron-Cohen & Robertson, 1995; Tager-Flusberg et al., 1997), which are problematic for a theory which states that both ToM and EF tasks use common logical structures, but that ToM tasks may require extra knowledge of mental states. It is also harder to see how FB explanation tasks may be framed in the embedded conditional framework, and yet



they still correlate highly with tests of EF requiring rule use (Hughes, 1998a), which implies that even the correlation between ToM and FB in TD 3-year-olds may depend on more than a common rule structure.

In addition, whilst this rule might define a similarity between some EF and ToM tasks, it is silent on the question of whether ToM deficits in real life are attributable solely to a failure to reason with embedded conditionals. It seems unlikely that all understanding of other minds, in TD children and children with ASD, can be reduced to a rule such as this. Whilst this theory may have some explanatory value for the correlation seen between the DCCS and FB tasks in TD children, it cannot explain either the relationship seen with all FB tasks, nor can it explain the deficits seen in ASD.

## **Summary**

There are four distinct possibilities that could explain the observed relationship between EF and ToM on a psychological level. After evaluating the available evidence, it is clear that there is no straightforward answer from the research to date. The field suffers perhaps from an overly narrow view of 'theory of mind', taking it often to mean whether a child passes or fails a FB tasks, and a lack of graded and comparably difficult EF and ToM tasks. However, it does appear to be clear that executive factors in FB tasks cannot account for the relationship seen between the tasks, and it seems unlikely that failing to grasp the embedded conditional can account for all the difficulties seen in ASD. Therefore we are left with the two functional dependency options, that ToM is a prerequisite for EF, or that EF is a prerequisite for ToM. These two theories cannot co-exist, particularly if we continue to take a narrow view of ToM and to measure it by the FB task.

The literature on children's ToM has long suffered from the tendency to categorise ToM as a unitary concept, measurable by the FB task, at the expense of all earlier and later developments in the understanding of mental states. It is only necessary to compare TD 3-year-olds to children with ASD to see a ToM cannot be reduced to passing or failing FB. Three-year-olds and children with ASD both fail FB tasks in the majority of cases. However, TD 3-year-olds are not at all like children with ASD – their apparent lack of FB understanding does not result in the profound difficulties in socialisation and communication exhibited by children with ASD. To equate their performance seems to miss out on the dramatic differences between their

understanding of minds and mental states. As Perner (1999) points out, FB tasks do not provide a privileged access to a child's ToM. They demonstrate only the existence of a certain ToM capacity, the understanding of mistaken belief. This may indicate a wider impairment in attributing thoughts and beliefs, as it appears to in ASD, or it may simply indicate that children are not yet able to attribute mistaken beliefs in the context of experimental tasks, as it appears to in young TD children. These young TD children demonstrate in many other ways that they are able to attribute a range of mental states and emotions (see chapter 1). These earlier precursors to a 'full' ToM may be crucial in the development of EF, but as yet there has been little attempt to assess their development.

Having said this, there may also be reason to postulate a special relationship between a representational ToM and executive functioning. The data certainly suggests that there may be a particularly strong relationship between the acquisition of FB understanding and performance on EF tasks (see Tables 4.1 and 4.2), which may not generalise to other, more advanced tests of ToM. Since, as argued above, it seems unlikely that this relationship is due to task demands, it appears that there may be a particular relationship between the acquisition of a representational understanding of mind and executive functioning. This may also be part of a more general relationship between the two, where early ToM precursors predict later EF and vice versa, but the evidence for a 'special relationship' between the ages of 3 and 5 in normal development is quite compelling. Perner's account (e.g. Perner & Lang, 2000) argues that meta-representation is in fact an integral part of executive functioning, that the ability to 'step back' from our actions and therefore monitor them effectively poses similar demands to the ability to step back from our thoughts or beliefs and therefore attribute other thoughts to other people. It is true that many of the early skills needed in the development of EF (e.g. self awareness, action-monitoring) are very close to those skills needed to consider the contents of one's own and other's minds, and Perner (1999) makes a convincing case for the need for meta-representation in inhibiting a prepotent schema. Russell's approach (e.g. Russell, 1997b) in contrast argues that self-monitoring is needed for meta-representation. This again is quite specific about the representational understanding of mind being linked to EF. Others have also suggested there may be a special relationship – Harris (1992) suggests that imaginative flexibility is necessary for a child to represent mistaken beliefs, although he does not link this to executive functioning.

It is less clear whether this relationship is also seen in those with developmental disorders, specifically those with ASD. There is some evidence for a relationship between ToM and EF in high-functioning individuals with ASD, but less for a relationship in lower-functioning individuals (e.g. Ozonoff et al., 1991; Zelazo et al., 2002). This may be because lower-functioning individuals with ASD often do not appear to acquire any form of a representational ToM. This would have two consequences, first, that the FB tests commonly used would show little or no range in performance and therefore a relationship could not be demonstrated for methodological reasons, and secondly, that if there is a special relationship between a representational understanding of mind and executive functioning (as suggested above), this would not be evident in a group with no representational understanding of mind. However, this group may be particularly interesting when it comes to examining the causal direction of the relationship between ToM and EF. If a representational understanding of mind is necessary for EF, then all those who lack this representational understanding should have poor EF. If they do not, this implies either that the two domains are in fact functionally independent in this group, or that executive functioning is necessary (but not sufficient) for a ToM. Training this group in either EF or ToM would help to elucidate the relationship still further, since if a representational ToM is necessary for executive functioning, then training children in ToM may improve their EF, and vice versa.

The next chapter describes a study looking at the relationship between ToM and EF in lower-functioning individuals with ASD and MLD – precisely those individuals who appear to lack a representational ToM. Chapter 6 will then review the literature on training ToM and EF, and chapter 7 will report the findings of a ToM and EF training study with lower-functioning children with ASD and MLD.

## **CHAPTER 5**

### **STUDY 2: ToM AND EF IN CHILDREN WITH ASD AND MLD**

#### **Introduction**

Following the discussion in chapter 4 on the relationship between EF and ToM, this study tests the relationship in lower-functioning children with ASD and MLD, all of whom failed at least 2/3 standard FB tasks.

The relationship between ToM and EF has attracted much interest in recent years. Whilst the relationship in young TD children has been fairly well elucidated (see Perner & Lang, 1999 for a review), far less research has been done with those with developmental disorders such as ASD. A wide range of studies have found deficits in participants with ASD on ToM tests, including first-order FB tasks (Baron-Cohen et al, 1985, 1986, Leekam and Perner, 1991, Perner et al, 1989, Swettenham, 1996, Swettenham et al, 1996) second order FB tasks (Baron-Cohen 1989, Bowler, 1992, Happe, 1993, Ozonoff et al, 1991), stories requiring mental state attributions (Happe, 1994) and tests assessing the ability to make mental state attributions from the eye region of the face (Baron-Cohen et al., 1997; Baron-Cohen, Wheelwright, Scahill, Spong, & Lawson, in press). Deficits have also been found in executive functioning (see Sergeant et al., 2002 for a review), particularly in set shifting and flexibility. The co-existence of these two deficits has led some authors to postulate that ToM and EF are in fact fundamentally related in ASD, to the extent that one causes the deficits seen in the other (e.g. Russell, 1997b; Wimmer, 1989). However, surprisingly few studies have directly tested the relationship between the two areas, and to date these studies have found associations only in higher-functioning individuals with ASD (Ozonoff et al., 1991; Zelazo et al., 2001). There is evidence from a single case study that it is possible for a child with ASD to have good levels of EF whilst still demonstrating deficits in ToM (Baron-Cohen & Robertson, 1995). This relationship may have strong implications for our understanding of autistic spectrum disorders, and the core deficits involved.

The present study examined ToM and EF in a fairly low functioning sample of children with ASD and MLD, all of whom were selected for having failed basic first order tests of FB. Therefore the MLD group had atypically low ToM performance (as

compared to other children with MLD), and those children with ASD who pass FB tasks (and who appear in virtually every study of ASD and ToM) have been excluded. Looking at this group may help us to distinguish whether the relationship between EF and ToM is specific to the transition to a representational understanding of mind (as measured by FB tests), or whether it is also evident in those who do not yet pass FB tasks. Including a control group with MLD is important in enabling us to distinguish between deficits which many children with developmental disorders share, and those that are more specific to ASD.

The performance of these children with MLD who fail FB tasks may be crucial to our understanding of both ToM and ASD. If we argue, as many researchers do, that a deficit in ToM is fundamental to ASD, and explains many aspects of autistic behaviour (e.g. Baron-Cohen, 2000), and we take the FB task as a marker of ToM, then why are those children with MLD who failed FB tasks not demonstrating autistic-like behaviour? Does FB failure mean something different in their case, and if so what? If children with MLD are failing FB tasks because of task-specific factors rather than due to genuine ToM deficits, they should be able to demonstrate their relative competence in other ToM tests and on real life measures. This study included 8 ToM tasks, and the prediction was that children with MLD would perform better than children with ASD overall, despite being selected on failing a FB unexpected transfer task. A questionnaire containing items for which a ToM is thought to be necessary (e.g. white lies, deception, recognising complex emotions) was also completed by the teachers of the participants, and the prediction was that the MLD group would perform better on this real-life measure of ToM skills than would the members of the ASD group. This will be discussed in chapter 9.

The other predictions for this study, based on the literature to date, were that the group with ASD would demonstrate deficits in both ToM and EF when compared to verbal age matched controls with MLD, even when both groups were selected due to their poor performance on FB tests, and that ToM and EF (set shifting and flexibility in particular) would be related in both groups (following Ozonoff et al., 1991; Zelazo et al., 1996).

## **Method**

### **Participants**

Participants were recruited from eight schools for children with special educational needs in the greater London area. They had all participated in Study 1, and were included on the basis of two criterion. First, that they had to have a VMA (as assessed by the BPVS II; Dunn et al., 1999) over 4 years 3 months (as it was thought that they needed at least this level of language to understand the training programmes for which they were being selected, and since TD children with this level of language would be expected to pass FB tasks), and secondly, that they must fail at least 2/3 FB tasks which involved predicting the contents of another's mind. To this end, an additional unexpected transfer FB task was administered to those children who failed only one of the two tasks (the 'Sally-David' task and the 'Smarties – other' task) administered in Study 1. This task was analogous to the Sally-David task described in Study 1, but used a toy bear and hedgehog with coloured boxes instead of the illustrated pictures used in the Sally-David task. The ages of the children ranged from 6 years 5 months to 15 years 3 months.

The children formed two groups. Twenty-one had non-autistic moderate learning difficulties (MLD), and 27 had autistic spectrum disorders (ASD). Twenty of the children in this group had a statement of special educational needs with ASD or autism as their primary diagnosis, one had a diagnosis of Asperger's Syndrome. The remaining 6 (who were from a single school and who were all being educated in classes for children with autistic spectrum disorders) were described on their records as having social and communication disorders. These children were assigned to the ASD group after completion of a checklist of symptoms based on DSM-IV after discussion with their teachers.

The relative ages and ability of each group are given in table 5.1. A oneway ANOVA found significant differences between the groups on Raven's Coloured Progressive Matrices (CPM) ( $F = 10.12$ ,  $df = 1,46$ ,  $p < .01$ ) and CA ( $F = 6.63$ ,  $df = 1,46$ ,  $p = .01$ ) only. This discrepancy in nonverbal ability will be taken account of where appropriate.

The tasks were administered by a single researcher in a quiet room in the child's school. All tasks were typically administered in a single session, lasting about

an hour. However, in the case of some of the younger children this was divided into two half hour sessions to maximize their concentration. The tasks were presented in one of two fixed orders, counterbalanced in the two clinical groups. There were no order effects on task performance.

**Table 5.1. Age and ability of the ASD and MLD groups: mean (s.d.)**

	ASD (n = 27)	MLD (n = 21)
CA (years)	10.35 (2.55)	12.02 (1.69)
BPVS VMA (years)	6.52 (1.75)	7.12 (1.93)
TROG VMA (years)	5.00 (.99)	5.50 (1.34)
Raven's CPM (raw scores)	22.93 (7.17)	17.14 (4.79)

The tasks described in this study formed the pre-test for the training study described in chapter 7, and were therefore designed with that in mind. In order to minimise the repetition of tasks in the training study, variations or subsections of some of the tasks were used. This means that the Eyes task, for example, was split into two sections, each containing 14 items. These sections were matched for difficulty using the results from a sample of 52 TD eight-year-olds (Fisher, Dunn and Hughes, submitted). Alternative forms of some of the other tests were used, for example a marble was used for the 'penny' hiding task, but no changes were made to the underlying structure of the tasks.

## **Measures**

### **General ability**

**Raven's Coloured Progressive Matrices (CPM) (Raven, Raven, & Court, 1998):** Non-verbal ability was assessed using the coloured set of Raven's Matrices, which are recommended for children and those with learning difficulties. The test was administered according to the test manual. As standardized scores are not available for the full age range tested here, raw scores were used.

**Test for Reception of Grammar (Bishop, 1989):** The administration and scoring of this test is described in Study 1. For ease of comparison across a wide age range the age equivalent score was used in this study. This will be referred to as the TROG VMA score.

**British Picture Vocabulary Scale (Dunn et al., 1999):** The administration and scoring of this test is described in Study 1. For ease of comparison across a wide age range the VMA score was used in this study.

## **ToM tasks**

**Sally-David:** This task is described in the methods section of Study 1.

**Smarties self and other:** This task, which involved two FB (FB) questions, is described in the methods section of Study 1.

**Extra unexpected transfer FB task:** This task was only administered to those who performed inconsistently on the Sally-David and the Smarties-other task. It used a toy bear 'Barney' and a toy hedgehog 'Henry', two coloured boxes and a ten pence piece. The child was told that Henry had ten pence, which he was going to place in one of the boxes to keep it safe whilst he went outside. Henry then left. Barney then entered and moved the coin into the other box. Henry then returned and the child was asked four questions:

FB question: Where will Henry look for his coin?

Justification: Why will he look there?

Reality control: Where is the coin really?

Memory control: Where was the coin first of all?

**Marble hiding Deception task** (based on Baron-Cohen, (1992): A marble was used in this task. The task was introduced by three trials of the researcher hiding the marble in one of her hands behind her back. She brought her hands forward and the child was invited to guess which hand the marble was in. After this, the child was asked to have a turn, with the instructions to 'See if you can trick me. Hide it really well, just like I did.' The child was given three trials, and was marked as having passed a trial if they successfully put both hands behind their backs, hid the marble in one hand, and brought both closed hands forwards, with the marble hidden, and did not tell the researcher where the coin was. They were scored as passing the whole task only if they passed all three of these trials.

**Seeing Leads to Knowing** (based on Pratt and Bryant, 1990): A small teddy bear, a toy hedgehog, a toy train and a box were used in this task. The toy train was placed inside the box. The two characters then entered the room. The bear looked inside the box, whilst the hedgehog touched the outside of the box. The child was then asked the following questions:

Seeing-Knowing (S-K) question: Who knows what is in the box?

Justification question: How does he know?

Children were credited with a pass if they gave the correct response to the S-K question.



Knowing/guessing (based on Kazak, 1997): This task was introduced by showing the child a sealed envelope. The child was told that the researcher had received this envelope in the post that morning, and hadn't looked inside it yet. The child was then shown the contents of the envelope whilst the researcher looked away, and said that she was not looking. The child was then asked the following questions, with the order of 'know' and 'guess' counterbalanced:

**Self knowledgeable:** Do you know what's inside the envelope, or do you have to guess?

**Other ignorant:** How about me, do I know what's inside the envelope, or do I have to guess?

Another envelope was then introduced, and the researcher looked inside without showing the child. They were then asked (with 'know' and 'guess' counterbalanced in order):

**Self ignorant:** Do you know what is inside the envelope, or do you have to guess?

**Other knowledgeable:** How about me, do I know what is inside the envelope, or do I have to guess?

This task was scored in two sections, 'self' and 'other'. Children were credited with passing only if they passed both questions in a section (e.g. self ignorant and self knowledgeable.)

Children's 'Reading the Mind in the Eyes' task (Baron-Cohen et al., 1997): Children were given half of the children's version of this task. This task involved the children being shown 14 pictures of someone's eyes. For each picture, they were read four words that described what the person in the picture might be thinking or feeling, and asked to choose one, either by pointing to the word on the page or by saying their choice. The items included were selected by comparing scores on this task from an earlier study of TD children (Fisher, Dunn, & Hughes, submitted), to create two halves on which scores were comparable. Potential scores therefore ranged from 0 – 14.

### **Non-ToM control task**

False photograph (Zaitchik, 1990). A Polaroid camera, a toy hedgehog ('Henry'), a box and a basket were used in this task.

The hedgehog was placed in the box, and a photo was taken. The photo was placed face down on the table, and the hedgehog was moved from the box into the basket. The child was then asked.

FP question: Where will Henry be on the photo?

Justification question: Why?

Memory control: Where was Henry first of all?

Children were given 1 mark for a correct answer to the FB question as long as they passed the control. All children passed the control question.

### **Executive Function**

**Card Sort:** This was a simplified version of the Wisconsin Card Sort Task. It was modified to make it shorter and easier for the present sample of children with intellectual impairments. A pack of 27 cards was used. The cards varied along 3 dimensions, colour, number and shape. Three cards were placed in front of the child, with a single red triangle, two blue circles and three yellow rectangles on them. The children were told that they were going to play a guessing game, and that the researcher was not allowed to tell them very much about how to play it, they had to work it out for themselves.

They were asked to match each of the cards in the pile to one of the three in front of them, and told there was a rule that they must work out to get the matching right. The researcher gave them immediate feedback on their performance. Once they had matched six consecutive cards correctly, the cards were reshuffled, handed back to them and they were told that there would now be a new rule for them to work out.

If the child did not work out the rule within 24 trials, they were told the rule and given 10 trials to demonstrate their ability to match by the given dimension. If they sorted more than 2 cards incorrectly within those 10 trials, the task was discontinued. Otherwise, the cards were reshuffled and they were asked to try and find a new rule.

If the child failed to work out two consecutive rules, the task was discontinued. There were four rules, presented in the following order; Colour – Shape – Number – Colour.

Scoring was as follows.

Five variables were calculated from this task.

**1. Number of categories achieved,** a general measure of success on the task, was calculated by counting the number of categories the child achieved without help. This could range from 0-4.

**2. Percentage conceptual level responses**, a measure of what percentage of the time the participant was sorting by the correct rule, was calculated as follows. All correct responses that occurred in runs of three or more were counted, and these were calculated as a percentage of total responses. Participants who scored highly on this measure were therefore achieving each category quickly with minimal trial and error.

**3. Percentage perseverative errors**, a measure of a child's tendency to remain 'stuck' on an earlier category of cards. This variable was calculated as follows.

A count was taken of perseverative responses, that is responses that would have been correct in the previous stage. They had to be either unambiguous (i.e. they were only matched by the perseverative category), or, if they were ambiguously matched, had to be contained within runs of 3 or more all sorted by the same category.

In the first category, the first unambiguous error became the 'perseverated to' category. The perseverated to principle could change within a single stage of the test if the participant made three unambiguous matches in succession according to another rule. Ambiguous responses were allowed between the unambiguous ones, as long as they were sorted according to the same rule as the unambiguous responses. The perseverative responses that were also errors were then counted, and this was expressed as a percentage of the total number of responses made over the course of the task.

In order to capture a different sort of perseveration, a score was also given for perseverative strategies other than matching to a category. These included placing all the cards in a single pile, or dealing the cards onto the three piles in order (e.g. from left to right), regardless of the feedback. A percentage error score was calculated as above. Very few children made such errors, and so in order to avoid multiple comparisons, this count was added to the percentage perseverative error score calculated above to create a single variable of percentage of perseverative errors.

**4. The number of trials to complete the first category** was counted as a measure of speed of understanding the nature of the task. This had a minimum of six and a maximum of 24, since if the child failed to get 6 correct within 24 responses they were told the rule.

**5. Failure to maintain set** was calculated as the number of times a participant made 4 correct responses but failed to get 6 right and therefore complete the category. This variable was included in order to identify those children who appeared to find a rule but then were unable to sort by it consistently.

Trails task (Reitan, 1958): This task has two sections. Trails A requires the child to connect a series of circles according to the numbers on them. The numbers go from 1-15, and there is a short practice first. The child is timed as they connect the circles. Trails B has both numbers and letters. The child must connect numbers and letters alternately, keeping both in order (i.e. 1-A-2-B). Again, there is a short practice, and they are timed. This task was discontinued if it was obvious that the child could not count or did not know the alphabet, or if they took over 300 seconds to complete either section of the task. If the child went wrong, they were told to go back to the last circle where they were right, and try again. Therefore, whilst they were not scored on their errors (although they had to complete the task accurately to be given a time) it added to their total time. A difference score, reflecting the difficulty of switching set, is then calculated by subtracting the time for Trails A from the time for Trails B.

Go-NoGo (Drewe, 1975): This computer based task tests inhibitory control. The child is told that they will see aeroplanes and bombs appearing on the screen, and they must shoot the aeroplanes by pressing a button on a mouse, but they must never shoot the bombs. No feedback was given for any of the trials. The task had 180 trials and lasted 5 minutes, 12 seconds. Thirty percent of trials were bombs, and the remainder were planes. The percentage commission errors and omission errors were calculated. Signal Detection theory was used to provide a measure of the participant's sensitivity to the task, and their response bias. A non-parametric measure of their sensitivity,  $A'$  was used. The Log-linear correction as proposed by Hautus and Lee (1998) was used to cope with response frequencies of zero.

## **Results**

### **Performance on ToM tasks**

The ToM tasks (apart from the Eyes task) and the False photograph task were all scored as pass/fail. The percentages of children passing each task, and mean scores for the Eyes task, are given in Table 5.2. Note that the children were selected for this study by their poor performance on the Sally-David and Smarties-other tasks, and so these scores are necessarily low.

Chi-square analyses showed that there were significant difference between the groups on the Smarties other ( $\chi^2 = 4.70$ ,  $df = 1$ ,  $p < .05$ ) and the Know-guess other ( $\chi^2 = 6.03$ ,  $df = 1$ ,  $p = .01$ ). There were trends towards significance on the Seeing leads to Knowing ( $\chi^2 = 2.90$ ,  $df = 1$ ,  $p = .09$ ), Know-guess self ( $\chi^2 = 2.90$ ,  $df = 1$ ,  $p = .09$ ), Smarties self ( $\chi^2 = 3.18$ ,  $df = 1$ ,  $p = .07$ ) and Sally-David ( $\chi^2 = 2.89$ ,  $df = 1$ ,  $p = .09$ ). This last is towards the ASD group performing better, all the others are in the opposite direction. Only the Marble Hiding task out of these ToM tasks does not approach significance, with both groups performing very poorly. There was no difference on the False photograph task. Performance on this task was much higher in both groups than performance on the supposedly equivalent FB unexpected transfer tasks. This was unsurprising in the case of the ASD group, but in the case of the MLD group the prediction had been that they would be as impaired on this task as they were on FB tasks. There was no group difference on the Eyes task.

**Table 5.2 ToM and control task performance split by group (percentage passing or mean; s.d)**

	ASD (n=27)	MLD (n=21)
Sally David (% pass)	22.2	4.8
Smarties – other (% pass)	14.8*	42.9*
Smarties – self (% pass)	40.7	66.7
Seeing -Knowing (% pass)	70.4	90.5
Marble Hiding (% pass)	7.4	14.3
Know-guess – other (% pass)	40.7**	76.2**
Know-guess – self (% pass)	29.6	57.1
False photo (% pass)	70.4	71.4
Average no. of ToM tasks passed (max = 7)	2.26 (1.26)	3.52 (1.60)
Eyes task (14 items)	4.30 (2.32)	5.10 (2.26)

\*  $p < .05$ , \*\*  $p < .01$

### **Agreement between ToM tasks**

In order to look at the level of consistency across performance on ToM tasks, and to look at the justification for creating a ToM aggregate, Cronbach's alpha was calculated for all eight ToM tasks with the groups separately. This gave a very low value of .16 for the MLD group and .28 for the ASD group. The Eyes task appears to be a very different type of task to the other ToM tasks used, and previous work by the author (Fisher et al., submitted) had found that it did not relate strongly to other ToM measures. In addition, it is a continuous measure whilst all the others are categorical.

Removing the Eyes task brought the value for the MLD group up to .62, whilst the ASD group alpha dropped slightly to .20. The Sally-David task was also removed from the aggregate. The rationale for removing this task was that it had effectively been the main task on which the groups had been selected - children were only included in this study if they failed at least one unexpected transfer task (either in conjunction with failing the Smarties-other task or an additional unexpected transfer task). Removal of the Sally-David task brought the MLD alpha up to .66, and the ASD group's alpha up to .33. The aggregate created therefore consisted of 6 items, all the dichotomous ToM tasks except for the Sally-David unexpected transfer task. This score (equivalent to no. of tasks passed) ranged from 0 to 6, with a mean of 2.67 (s.d. = 1.58). The two groups had a highly significant difference on this aggregate (ASD mean = 2.04, s.d. = 1.26, MLD mean = 3.48, s.d.=1.60;  $F(1,46) = 12.21$ ,  $p = .001$ ). This aggregate did not correlate significantly with the Eyes task in either group (ASD  $r = .31$ ,  $p = .12$ , MLD  $r = -.08$ ,  $p = .72$ ).

### **Performance on EF measures**

Table 5.3 gives the mean group scores for the EF tasks administered.

***Table 5.3 EF task performance split by group.***

<b>Card Sort</b>	<b>ASD</b>	<b>MLD</b>
% perseverative errors	38.23 (23.14)	26.84 (18.15)
Number of trials to complete first category (max = 24) <sup>a</sup>	9.92 (5.21)	10.00 (4.32)
Number of times failed to maintain set	0.19* (0.40)	0.52 * (0.68)
No. of categories achieved (max = 4)	1.96 (1.45)	1.81 (1.29)
% conceptual level sorting.	28.01 (23.32)	30.03 (20.84)
<b>Trails<sup>b</sup></b>		
Trails A (secs)	38.99 (17.09)	32.31 (8.36)
Trails B (secs)	94.78 (54.70)	88.42 (47.67)
Trails Difference (secs)	55.80 (42.42)	56.12 (45.19)
<b>Go-NoGo<sup>c</sup></b>		
% commission errors	41.77 (22.57)	41.43 (22.69)
% omission errors	22.24 (18.09)	20.04 (17.41)
A'	0.76	0.78

<sup>a</sup> 7 children with ASD and 5 children with MLD did not complete the first category within 24 turns and so did not have a score for this variable

<sup>b</sup> 6 children with ASD and 5 children with MLD could not complete the second half of the Trails task within 300 seconds. These children are not included in the scores.

<sup>c</sup> Due to a malfunctioning computer the results for 10 of the children with ASD were lost for this task. Therefore these results represent only 17 children with ASD.

\* $p < .05$

A oneway ANOVA found that the only significant difference on the Card Sort task was on the 'failure to maintain set' variable ( $F(1,46) = 4.68, p < .05$ ) on which the MLD group was worse than the ASD group. The difference on the 'percentage perseverative responses' approached significance ( $F(1,46) = 3.44, p = .07$ ), with the MLD group outperforming the ASD group. No other variables in table 4.2 were significantly different between groups (all  $p$ -values  $> .29$ ). The two groups differed on non-verbal ability, and therefore this might have accounted for the difference on the 'failure to maintain set' variable, and indeed the lack of differences between the other variables. Since Miller and Chapman (2001) suggest that it is theoretically suspect to co-vary for factors that differ between groups in a case such as this, forced entry hierarchical multiple regression was done as an alternative. Failure to maintain set was entered as the outcome variable, and Raven's matrices were entered in the first block ( $R^2 = .002, F = .08, p = .77$ ). In the second block, group (ASD or MLD) was entered ( $R^2 = .13, F = 3.13, p < .05$ ). Group was a significant predictor of performance over and above Raven's Matrices, which was in fact not a significant predictor alone or with group. Therefore it seems that the difference in non-verbal ability does not account for the group difference on failing to maintain set.

This analysis was repeated with the other EF variables in order to analyse the effect of non-verbal ability separately from the effect of group.

With percentage perseverative errors, in the first block, Raven's was not a significant predictor ( $R^2 = .05, F = 2.50, p = .12$ ). When group was added in the second block, it predicted perseverative errors along with Raven's Matrices ( $R^2 = .21, F = 5.94, p < .01$ ). Both Raven's and group made independent contributions to this model (Raven's  $t = -2.82, p < .01$ , group  $t = 2.99, p < .01$ ). With percentage conceptual level processing, Raven's was a significant predictor alone ( $R^2 = .15, F = 8.33, p < .01$ ), but group was a predictor over and above Raven's ( $R^2 = .25, F = 7.44, p < .01$ ). Both Raven's and group remained significant predictors in this model ( $p < .01$  in both cases). Neither Raven's nor group were significant predictors of the number of trials to complete first category (both  $R^2 < .02, F_s < 0.56, p_s > .46$ ). In the case of the number of categories achieved, Raven's was a significant predictor ( $R^2 = .19, F = 11.08, p < .01$ ), but group did not predict number of categories over and above it ( $t = 1.10, p = .28$ ).

In the case of Go-NoGo, neither Raven's nor group were significant predictors (both  $R^2 < .02, F$ -values  $< .37, p$ -values  $> .69$ ). In the case of the Trails, Raven's was a

significant predictor alone ( $R^2 = .18$ ,  $F = 8.31$ ,  $p < .01$ ), and group did not predict performance over and above Raven's ( $t = -0.44$ ,  $p = .66$ ).

Therefore, it seems that group predicted performance over and above non-verbal ability on the percentage of perseverative errors, percentage of conceptual level sorting (in both cases with the ASD group performing worse than the MLD group) and on failing to maintain set (with the MLD group performing worse than the ASD group).

The EF measures were simplified down to three variables in order to look at the relationships between EF and ToM. Within the Card Sort task, all variables correlated significantly or approached significance in the ASD group except for the Failure to Maintain Set variable. In the MLD group, all variables correlated significantly or approached significance except for the 'number of trials to complete first category' variable. Due to this differing pattern of results in the MLD and ASD groups, it was felt to be important to keep in all the measures rather than selecting one or two of the variables as the most informative measures. To simplify comparisons a 'Card Sort' aggregate was created combining scores from all 5 variables as follows. Each child was given a score of 0-2 for each variable. Those who were in the top 25% relative to the whole group were given 2, those in the middle 50% were given 1, and those in the bottom 25% were given 0. These scores were then summed to make a Card Sort aggregate score. This ranged from 0 – 10, with a mean of 4.58 (s.d. = 2.45). The MLD and ASD groups did not differ on this score (ASD mean = 4.85 (2.81), MLD mean = 4.24 (1.89),  $F = .74$ ,  $df = 1,46$ ,  $p = .40$ ). The Difference score was used for the Trails, and A' for the Go-NoGo.

There were no significant correlations between performance on the Trails task and the Card Sort aggregate, or between Go-NoGo and Trails or the Card Sort aggregate. Table 5.4 gives these correlations.

**Table 5.4 Pearson's correlations between EF tasks. ASD group (N = 27) above diagonal in bold, MLD group (N=21) below diagonal**

	Go-NoGo A'	Trails difference	Card sort aggregate
Go-NoGo (A') (high = good performance)		<b>-.18</b>	<b>.17</b>
Trails difference (high = poor performance)	.08		<b>-.12</b>
Card sort aggregate (high = good performance)	-.12	-.37	



### **Ability, age, ToM and EF.**

The relationship between ability, age and the ToM and EF measures was investigated as ability and age might account for any relationship seen between ToM and EF. Tables 5.5 and 5.6 report the correlations between the age, ability measures, the ToM measures and EF measures. To minimise the risk of a Type 1 error, the aggregates rather than individual task performance are used.

The Eyes task and the Card Sort were most closely related to ability, both verbal and non-verbal, in both groups. The ToM aggregate appeared to be less strongly related to verbal ability in the MLD group than in the ASD group, although this is based on effect sizes only, since no correlations were significant in either group. The Trails was related to ability in both groups, but this was stronger in the MLD group (although this again is based on effect sizes since the numbers for this task were small and the correlations are non-significant). The Go-NoGo was related strongly to age in the MLD group, there was no relationship at all between age and Go-NoGo in the ASD group.

***Table 5.5. Correlations between age, ability, ToM and EF for the ASD group (N = 27)***

Task	CA (years)	BPVS VMA	TROG VMA	Ravens
ToM aggregate	.11	.26	.37	-.13
Eyes task	.27	.23	.48*	.18
Card sort aggregate	.04	.39*	.44*	.52**
Trails (n=21)	-.27	-.28	-.03	-.38
Go-NoGo A'	.02	.19	.13	.00

\* p <.05, \*\*p<.01

***Table 5.6. Correlations between age, ability, ToM and EF for the MLD group (N = 21)***

Task	CA (years)	BPVS VMA	TROG VMA	Ravens
ToM aggregate	.00	.14	.14	-.07
Eyes task	.51*	.59**	.39	.54*
Card sort aggregate	.13	.27	.56**	.13
Trails (n=16)	-.33	-.45	-.48	-.34
Go-NoGo A'	.55*	.22	.05	.26

\*p<.05, \*\*p<.01

### **EF and ToM**

In order to look at the relationship between the ToM aggregate and EF measures, correlations were calculated within each diagnostic group. Table 5.7 gives these correlations.

The ToM aggregate did not correlate significantly with any of the EF measures in either the ASD or the MLD group (all  $p$  values  $> .45$ ).

**Table 5.7 Pearson's correlations between the ToM and EF measure for the ASD ( $N = 27$ ) and MLD ( $N = 21$ ) groups (ASD above diagonal in bold).**

	ToM aggregate	Eyes task	Card Sort aggregate	Trails	Go-NoGo A'
ToM aggregate		<b>.31</b>	<b>.07</b>	<b>.00</b>	<b>.20</b>
Eyes	-.08		<b>.38*</b>	<b>-.19</b>	<b>.15</b>
Card Sort	-.11	.40		<b>-.12</b>	<b>.17</b>
Trails	.01	-.45	-.37		<b>-.18</b>
Go-NoGo A'	-.12	.28	-.12	.08	

\*  $p < .05$

### **The Eyes task**

The relationship between the Eyes task and the EF measures was calculated separately as the Eyes task was not included in the ToM aggregate. The Eyes task correlated highly with EF tasks in both the ASD and MLD groups. However, when the TROG (which correlated with both Card Sort aggregate, Trails and the Eyes task) was partialled out all these correlations that were significant or had a trend towards significance became non-significant (Card Sort: ASD  $r = .22$ ,  $p = .29$ , MLD  $r = .24$ ,  $p = .31$ , Trails MLD  $r = -.27$ ,  $p = .30$ ), indicating that it may have been the relationship of both tasks to the TROG which accounted for their association rather than a genuine link between EF and the Eyes task.

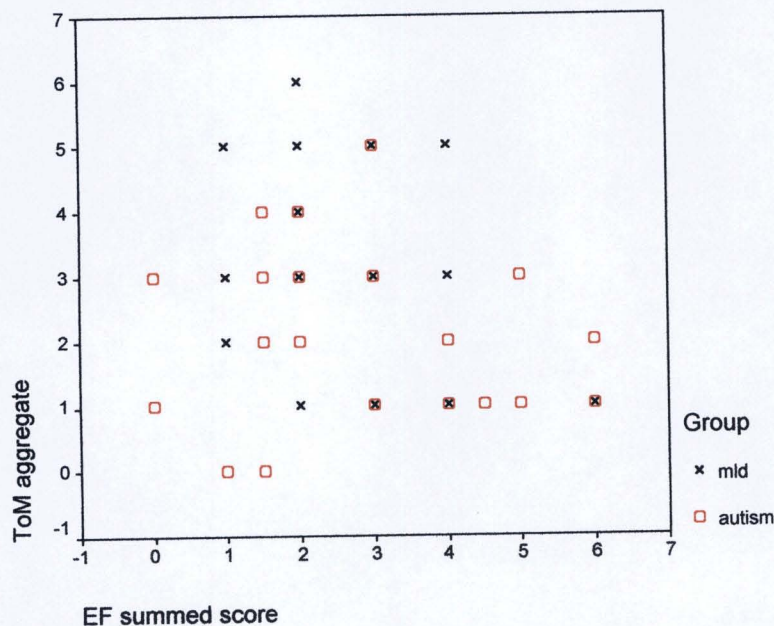
### **Individual patterns of performance**

If EF is absolutely necessary for ToM development, or if EF is necessary for the actual execution of ToM online (or indeed vice versa), it should not be possible to find children with good EF who have poor ToM (or vice versa). In order to look at this, an EF summed score was created by dividing all three EF variables into 3 bands, with the top 25% relative to the whole group getting a score of 2, the middle 50% getting 1, and the bottom 25% getting zero. These variables were then summed. Children who scored highly on this summed score had therefore had done well on all 3 EF tasks. This is not an EF aggregate and has not been used as such due to the low correlations between performance on the different EF tasks. For those 10 children

who had no score for the Go-NoGo task, the sum of their scores for the other two tasks was multiplied by 1.5 to give a pro-rated score.

As can be seen from the scatter plot in figure 5.1, there was no clear relationship between EF and ToM in either group. Overall the MLD group tended to do better on the ToM aggregate. There were individuals – particularly in the ASD group - who did very well on the EF tasks whilst doing poorly on the ToM aggregate. There were 4 individuals who had an EF score of at least 4 points higher than their ToM score. Three of these were in the ASD group (1 girl, 2 boys) and one boy was in the MLD group. Inspection of these individuals' age and ability suggested that the only obvious difference from the rest of the group was the high non-verbal ability of the MLD participant.

**Figure 5.1** Scatter plot of ToM v. EF, with markers set by group



There are also those individuals who did very well on the ToM aggregate whilst doing poorly on the EF aggregate. These individuals only existed in the MLD group, their existence in the ASD group was precluded by the lack of individuals in the ASD group who did well on the ToM aggregate.

There were 2 individuals who had a ToM score of at least 4 points higher than their EF score, and these were both boys in the MLD group. Their verbal and non-verbal ability appeared to be comparable to that of the rest, and if anything was slightly lower than that of the whole group.

### **Summary of the relationships seen between age, ability, ToM and EF.**

There is no evidence for a relationship between ToM and EF in the ASD group. In the MLD group a single EF variable (failure to maintain set in the Card Sort) correlated with the ToM aggregate. A relationship between the Eyes task and EF disappeared when the TROG was co-varied out. Individuals in both the MLD and ASD group existed who had very good EF but poor ToM, and individuals existed in the MLD group who had relatively good ToM but poor EF. There were no individuals in the ASD group with good ToM performance but poor EF, but this may be because there were very few individuals in the ASD group with good ToM performance.

## **Discussion**

### **Theory of Mind and Executive Function in children with ASD and MLD**

A hypothesis of this study was that, whilst children with MLD can be found who fail FB tasks, these children would still perform better on a range of ToM tasks than would children with ASD. The children in this sample were chosen because they failed at least 2/3 standard FB tasks. The present study indicates that whilst a proportion of children with MLD fail FB tasks, these children still demonstrated superior mentalising abilities in other experimental ToM tasks when compared to children with ASD. These results suggest that whilst some children with MLD might have difficulties with standard FB tasks, this should not be taken as an indication that they have comparable socio-cognitive difficulties to those with ASD. Study 1 of this thesis (chapter 3) indicated that children with MLD might be particularly vulnerable to task-specific demands of ToM tasks, and that therefore care should be taken when comparing them with children with ASD on FB. However, there is some indication in this study that task specific demands cannot account for all of the variability seen in ToM performance in children with MLD – it is intriguing that the ToM tasks are *more* strongly related in the MLD group than in the ASD group. It was also the case that, contrary to predictions, the MLD group had no difficulties with the ‘false photograph’ task, a non-social task designed to mirror the representational demands of the standard FB task without reference to mental states. Both the ASD and MLD groups found the false photograph task relatively easy, with over 70% passing in both groups. This

indicates that whatever it is about the FB task that the MLD group found hard is not a component of the False photograph task. This is in contrast to TD 3-year-olds, who have been reported to find this task as hard as the FB task (Leekam & Perner, 1991), although other studies (Slaughter, 1998) have found that the false photograph task is easier for young TD children than the FB task. It is possible that the False Photograph task is easier than a FB task by virtue of children's frequent exposure to photographs and their properties – which may effectively provide repeated practice on this type of task.

The poor association between ToM tasks seen in the ASD group may reflect the fact that three of the tasks in this study differed from more standard ToM tasks in that the failure to understand mental states will not lead to a consistently incorrect response. Whilst in the standard unexpected transfer task, a child who does not appreciate that a mental state can differ from reality will consistently say that the protagonist will look in the wrong box, in tasks such as Seeing Leads to Knowing a child without ToM would be expected to perform at chance when given a choice between two dolls who have both had some contact with the box. This will lead to essentially random performance in those with difficulties in ToM, and will therefore lead to poor agreement between tasks. This may account for the poor inter-task agreement seen in the ASD group. By contrast, children who understand the character's mental states will perform correctly above chance, hence, perhaps, the superior alpha for the ToM tasks in the MLD group.

The relationship between language and ToM has been discussed elsewhere (chapters 2 and 3). Study 1 found that grammar was a significant predictor of performance on standard FB tasks over and above vocabulary and age in ASD, but not in MLD. The present study found a trend towards a relationship between grammar and ToM in the ASD group only, despite the ToM tasks used in this study covering a broader range than the standard unexpected transfer and deceptive box tasks used in the earlier study. The lack of an association in the MLD group indicates again that there may be a special relationship between language and ToM in ASD, (in line with study 1 of this thesis and Happé; 1995).

There were some EF differences between the groups, in line with earlier studies (e.g. Hughes et al., 1994; Ozonoff & McEvoy, 1994; Ozonoff et al., 1991; Rumsey & Hamburger, 1988). There was a trend towards the ASD group showing more perseveration on the card sort task than did the MLD group, and this was

significant once the effects of non-verbal ability were accounted for. The ASD group also showed less conceptual level sorting once their superior non-verbal ability was accounted for – not surprisingly, since perseverative errors and conceptual level sorting are closely inversely related. However, there were also differences from earlier studies (e.g. Ozonoff et al., 1991) in that the ASD group in this study were *better* at maintaining set on the Card Sort tasks than were the MLD group – this is a measure of how many times individuals sorted four cards correctly in a row but then failed to continue to get the six necessary to complete the category. There were also no differences on other measures of EF, the Trails task and Go-NoGo, in contrast to Rumsey and Hamburger (1988) who found differences on the Trails task when comparing adults with ASD with normal adults, and Ozonoff et al (1994) who found differences between an ASD group, a group with Tourette's Syndrome and a TD group on the Go-NoGo task. This may reflect the nature of the present control sample, all of whom had MLD, in contrast to the normal IQ controls used in the Rumsey and Hamburger and Ozonoff et al studies. It is possible that some of the EF deficits seen in ASD may be evident only in comparisons with controls who do not have learning difficulties, and may in fact be shared with other developmental disorders. However, this cannot explain all the differences since the control group in Ozonoff's (1991) study had learning difficulties and were still significantly better on an EF aggregate and on separate variables from the WCST and Tower of Hanoi.

### **The relationship between ToM and EF**

A prediction of this study was that the EF tasks would be related to performance on ToM in children with ASD and MLD, reflecting the findings from earlier studies (Ozonoff et al., 1991; Zelazo et al., 2001). This was not found to be the case, in contrast to the earlier studies cited above. What might explain this difference? Task differences are one candidate - all other studies have used the standard unexpected transfer FB task, which was excluded from the ToM aggregate in this study as the children were effectively selected by their failure on it. As suggested in chapter 4, there may be a special relationship between the acquisition of a representational ToM (as assessed by a FB task) and EF, and that therefore the relationship is far less evident in a study which includes only children who appear to lack this representational understanding. It also could be that children who pass these standard FB tasks demonstrate superior EF than do those who fail them, and that this

dichotomous distribution accounts for the observed relationship. It could also be because of floor effects on FB tasks in these groups – the sample in the present study represents a group who were near floor on standard FB, however, due to the inclusion of other ToM tasks a range of ability was still found to be present. If the relationship between ToM and EF in ASD only holds in higher functioning individuals – or possibly only in individuals who pass FB tasks, this may be because higher-functioning individuals with ASD are using cognitive strategies to solve FB tasks which themselves involve a degree of EF – for example, the language based strategies which were suggested in chapter 2 would require a degree of ‘holding in mind’ and shifting between perspectives which may be shared with EF tasks such as the Card Sort or tests of working memory. Another possible explanation for the lack of an association is that the performance on the ToM aggregate in the ASD group does not reflect genuine differences in ToM ability – it is argued above that performance on some of these tasks may be due to chance, and that even when ToM skills were lacking, consistent failure would not have been expected.

The relationship between the Eyes task and EF measures was unexpected. The Eyes task was significantly related or had a trend towards being related to the Card Sort task in both the MLD and ASD group. However, this became non-significant when grammatical ability was partialled out. The lack of a relationship between the Eyes task and the ToM aggregate suggest that these tasks measure quite different elements of the understanding of mind – although both measure elements that individuals with ASD appear to have difficulties with. The Eyes task has been used mostly in high functioning individuals with ASD or Asperger’s Syndrome in the past, where a deficit has been found when compared to normal controls (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Baron-Cohen et al., in press). Quite why grammatical ability (as measured by the TROG) should be related to the Eyes task is a puzzle – there are no obvious grammatical demands in the Eyes task, and it is hard to see how grammar might be important in the development of emotion recognition from eyes, in contrast to how it might be important to the development of FB understanding.

There has been some debate over whether individuals with ASD consistently demonstrate either ToM or EF deficits, and what this implies for our understanding of the disorder (Baron-Cohen & Robertson, 1995; Russell, 1997b). There was no real evidence for ASD –specific EF deficits in this study, the ASD group only differed



significantly from the MLD group on one EF measure, and several of the highest scorers on the EF summed score were in the ASD group. However, this score was relative to the rest of the sample rather than an absolute score, and it is possible that both groups were performing poorly on EF. Since we had no TD control group we cannot rule this out. In contrast, the ASD group performed much worse than the MLD group on the ToM aggregate. There were individuals, both with ASD and MLD, who performed very badly on ToM whilst displaying very good EF. There were only individuals with MLD with the opposite pattern, but this may be due to the poor performance of many children with ASD on the ToM tasks. We should not draw too many conclusions from individual cases but these may indicate, along with Baron-Cohen and Robertson (1995) and Tager-Flusberg, Sullivan, & Boshart (1997), that EF and ToM may be functionally independent in lower-functioning people with ASD or MLD.

### **Conclusion**

This study found that children with ASD demonstrated ToM difficulties when compared to children with MLD even when all children were selected by their poor performance on standard FB tasks. It therefore seems that, whilst some children with MLD do have difficulties on standard FB, this does not mean that the majority of those who fail FB have ToM problems in the same way as do children with ASD. However, further work is needed to elucidate the nature of these difficulties in children with MLD, as this study indicates that task-specific factors do not account for all variation in performance, and performance on ToM tasks in MLD was unrelated to age, ability and EF.

This study also found that whilst children with ASD made marginally more perseverative errors than children with MLD on a card sort task, they were unimpaired relative to the MLD group on other EF measures (inhibitory control and set shifting). Performance on ToM tasks was unrelated to performance on EF in both ASD and MLD. This contrast with previous studies may reflect the lower-functioning nature of this sample, and the fact that the unexpected transfer FB task was not included in the ToM aggregate. These results tentatively suggest that it is possible to achieve (relatively) high levels of EF without a good ToM, and vice versa. However, this does not mean that improving EF might not improve ToM (or vice versa). Whilst good EF might not be necessary for ToM, training children on EF may



still be sufficient to improve their ToM or vice versa. In order to investigate this relationship further, a training study of EF and ToM was carried out (Chapter 7). Before reporting these results, a literature review of studies attempting to train either ToM or EF will be presented (Chapter 6).

## CHAPTER 6

### TOM AND EF INTERVENTION STUDIES

#### Introduction

As discussed in chapter 4, the relationship between ToM and EF is one which has recently stimulated much interest (e.g. Frye, 1999; Hughes, 1998a; Tager-Flusberg et al., 1997). Several studies have shown a direct relationship between the development of ToM and increasing executive control in normal development (see Perner & Lang, 1999 for a review). Children with ASD and other developmental disorders have been shown to have executive problems alongside their difficulties in ToM; (Ozonoff & McEvoy, 1994; Ozonoff et al., 1991; Pennington & Ozonoff, 1996), and there has been much debate over which of these deficits might be primary, and therefore causal to the other. Whilst a range of studies have found correlations between children's concurrent performance on EF and ToM tasks (e.g. Hughes, 1998a; Pennington & Ozonoff, 1996), only a very few have taken a longitudinal approach, and have therefore been able to look at possible causal relationships between the domains (e.g. Hughes, 1998b; Ozonoff & McEvoy, 1994). There is some evidence that early executive performance may predict later ToM performance (Hughes, 1998b), but other studies suggest that the relationship may go in the opposite direction (see Perner & Lang, 1999). It is also important to bear in mind that a ToM is only one milestone of a young child's developing social understanding, and failing to measure early antecedents of ToM might lead us erroneously to the conclusion that a certain level of executive control precedes the development of understanding of mind, when in fact it only precedes the understanding of mistaken beliefs. It is possible that early antecedents of ToM could lead to the development of certain executive skills in TD children, which are important in later advances in mentalising.

Several different developmental accounts of the relationship between ToM and EF are possible. These have been discussed in detail in Chapter 4. Briefly, one option is that the relationship is a side effect of the tasks used to assess ToM. It has been suggested that children may fail tasks designed to assess ToM due to task demands such as flexibility or inhibitory control, rather than due to difficulties in

mentalising (Hughes & Russell, 1993). In this case, difficulties in one area may be masking competence in another area, and the association seen would be due to common task demands. Secondly, it is possible that a minimum competence in, for example, executive control is necessary for the development of mentalising skills, and that whilst the children are not failing FB tasks due to the executive demands of the task, they are failing due to limitations in the executive skills required either to develop a ToM, or actually use a ToM online (Hughes, 1998b). The converse could also be the case – it is possible that a certain understanding of mental states is necessary in order to perform well on EF tasks. Perner (e.g. Perner & Lang, 1999) proposed that an understanding of the representational nature of mental states (and an understanding of the causal relationship between mental states and action) is necessary in order to monitor behaviour sufficiently to pass EF tasks. Lastly, a third factor (e.g. embedded rule structure), may underlie both EF and ToM.

The relationship between EF and ToM in ASD is of particular interest because of the difficulties that children with ASD often demonstrate in both areas (although not always, see Chapter 5, this thesis), and the profound disabilities that these impairments may cause. If we could identify any causal relationship between the domains, we might be able to intervene at an early stage in order to facilitate development.

Given the difficulties of longitudinal studies, an alternative approach to testing the direction of the relationship between ToM and executive functioning would be to train children in either ToM or EF, and to examine the inter-relationships between the two domains. If, for example, children are failing ToM tasks due to their problems in inhibitory control and flexibility, rather than difficulties in mentalising, then training them in flexibility and inhibitory control might enable them to pass ToM tasks. Conversely, if they are failing tests of EF due to the representational demands of the tasks, then training them in ToM might enable them to overcome these difficulties and improve their EF performance.

This approach may also help us to distinguish between the difficulties seen in different groups of children. It may be that children with MLD are failing FB tasks for reasons other than deficits in ToM. As discussed in chapters 1,2 and 4, executive control and language ability are plausible candidates for limiting factors on ToM task success. Training these children alongside children with ASD may help us to identify

the differing sources of task difficulty in these groups and test the specificity of the ASD ToM deficit.

The success of some children with ASD on FB tasks is also surprising, and suggests that some children with ASD can successfully use and develop strategies for thinking about thoughts, at least in an experimental situation. The performance of children with ASD on FB tasks is strongly linked to verbal ability, and it has been suggested that more verbally able children may be able to compensate for some of their difficulties by using their superior language skills either to work out logically a non-ToM based method of passing FB tests (Happé, 1995), or as an alternative route to some genuinely better understanding of mind (Tager-Flusberg, 2002). It therefore seems possible that children with ASD might be able to be taught a verbally mediated strategy as a compensatory device for ToM.

The first step in designing a study of this kind is to devise training programmes in ToM and EF. Whilst an increasing number of studies have been done which aim to improve the performance of children with ASD on FB tasks (e.g. Hadwin et al., 1996; McGregor, Whiten, & Blackburn, 1998a; Swettenham, 1996; Swettenham et al., 1996) only a single study to date has aimed to improve the EF of individuals with ASD (Shimmon and Lewis, 2000). The majority of the studies that attempt to train any kind of EF have aimed to improve attention in children with ADHD. In the following review of intervention studies, the widest possible approach has been taken in order to identify any strategies that might enable children with ASD and learning difficulties access to the concepts involved. When examining ToM training studies, studies have therefore been included that have introduced modifications to FB tasks with the aim of improving the children's performance, as well as those that have set out to train children in a specific strategy. The review of EF training studies does not attempt to be exhaustive; the single training study with children with ASD is considered, and some representative studies with children with ADHD are considered insofar as they may inform the design of a programme for children with ASD.

### **Theory of Mind intervention studies.**

Several studies have set out to improve the ability of children to pass FB tasks, with varying degrees of success. The participants in this type of training study have mostly been children with ASD, but some participants with learning difficulties and

young TD children have also been included. These studies range from small scale intervention studies done on a one-to-one basis (e.g. McGregor et al., 1998a; Swettenham et al., 1996), to studies where groups of children or young people meet with the trainer for several months (e.g. Ozonoff & Miller, 1995). In the one-to-one intervention studies the child is trained, usually in a specific FB scenario, using different techniques to highlight problem areas. These techniques have ranged from feedback from a computer (Swettenham, 1996), to enhanced cues within the task (such as expressions of surprise on the part of the protagonist; (e.g. Bowler & Strom, 1998; Charman & Lynggaard, 1998; Parsons & Mitchell, 1999) to direct instruction in specific strategies. For example, Swettenham et al. (1996) used a 'photo in the head' technique as a cognitive prothesis to help the children think about thoughts, whilst Wellman, Baron-Cohen, Caswell et al. (2001) used thought bubbles in a similar way. The techniques used in some of these studies are compared in greater detail below. Initial discussion will focus on those studies that introduced modifications into repeated FB tasks, considering what they may tell us about the tasks and children's approach to them. The discussion will then move on to review the longer term intervention studies.

### **ToM task modifications.**

There are a number of modifications that may be introduced to the standard FB task. These modifications sometimes appear to help TD children younger than 4 to pass the tests, although children rarely pass at ages younger than 3:6 (Wellman, Cross et al., 2001). These modifications include clarifying the question (e.g. asking 'where will Sally look *first* for her marble?'), enhancing the salience of the representation by introducing picture cues, introducing a deceptive motive to the task, the child actively participating in the transformation, or reducing the salience of reality in some way, for example by not having a real object present. Several of these modifications have been used with children with ASD, and it is these that I will concentrate on here.

Modifications to a FB task often appear to improve performance, but whether this performance is transferred to tasks without cues is less clear. These studies generally introduce the cue in the course of a single session, and in some cases do not conduct post-testing without the cue - the aim is to see whether the cue can improve performance, rather than whether the children can pass the tests independently of the

cue. Studies will only be discussed where the modifications introduced inform our interpretation of the training studies which I will refer to later. Some of these studies use similar techniques to the training studies (for example, a photographic cuing technique), whilst others use repeated FB tests. Both of these are elements of training studies, and considering studies with these modifications but without training could help to clarify which elements of the training programmes may be helpful.

### **Behavioural and emotional cues**

Bowler and Strom (1998) used six repetitions of the unexpected transfer FB task with the inclusion of enhanced behavioural and emotional cues to the protagonist's FB. Their participants were children with ASD (9 in the experimental group, 8 controls), TD young children (15 older and 15 younger in the experimental group, 10 older in the control group), and children with MLD (8 in the experimental group only), all of whom failed an initial FB task. The cues used were very explicit - in the 'Surprise' form of the task, the protagonist returned, looked in the original hiding place, looked surprised, and gasped 'Gosh, my (object) isn't here!', whilst in the 'Action' condition the protagonist returned and looked in the original hiding place without saying anything. The child was then asked 'where does X think her (object) is?' No feedback was given after the child's responses. Only the children with ASD and the older TD children (CA mean 3:11) benefited from the cues, whilst those with learning difficulties and the younger normal children (CA mean 3:5) showed no improvement at all. The control groups of children with ASD (who were given 6 standard FB tasks and one own FB task in place of the cued FB tasks) also showed some improvement, with 4/8 passing some of the later tasks (in contrast to 0/8 in the first task). The most helpful version was the 'Surprise' condition - which had the most explicit cues - and which 8/9 of the children with ASD in the experimental group passed the second time it was administered. The improvement in the older normal children was much less dramatic, with only 5/15 passing the second 'Surprise' condition. A standard task was not administered last, and so it is not possible to tell if this improvement generalised to non-enhanced FB tasks. The children with ASD did improve on their second administration of a standard task (with 5/9 passing), however, since 4/8 of the control group with ASD also passed this task, it appears that the cues themselves may not be responsible for this improvement in performance, but

instead other more general factors may come into play (for example familiarity with the scenario and the experimenters).

These findings are unexpected, as they seem to indicate that children with ASD are particularly open to learning from enhancements of this type and from practice of the scenario, as compared to children with learning difficulties and young TD children. Given that the majority of children with ASD appear to fail to learn about false beliefs from the social experiences and cues they see in real life, whilst all TD children and most of those with learning difficulties generally do so, this is surprising. However, the cues used in this case were very explicit and there was no evidence to suggest that the children were able to use this experience to improve their performance on other FB tasks. Indeed, the 'Surprise' condition is scarcely a *false* belief task, since the protagonists explicitly stated their surprise at not finding the object before the child is asked the question, and therefore all the child needs to understand to pass is that saying something is indicative of what you think. As the authors suggest, it may be that the young TD group are simply too young to comprehend the tasks properly - previous studies have found it impossible to improve the performance on FB tasks of children aged below 3 years 7 months (Sullivan & Winner, 1991). The performance of the group with learning difficulties is puzzling - one suggestion was that the IQs of the children might have been important, rather than simply their VMAs. The group with learning difficulties were on average older than the group with ASD, and were therefore of lower IQ, as a consequence of being matched on VMA.

How should these results be interpreted? It is possible that the reasons why children with ASD fail these tasks – which may be genuine difficulties in considering mental states, or problems with inhibiting a prepotent response – are solved by the cues presented in the Bowler and Strom study. However, TD children (particularly young ones) and those with learning difficulties may be failing for alternative reasons, which are not solved by the cues. These alternative reasons are not obvious – but possibilities could be misunderstanding the question or even in this case failing to attend to the additional cues (particularly after 6 administrations of variants of the task, when the test questions must have become quite predictable). However, there is no obvious reason why this should be more marked in the non-autistic groups. Still, this study does raise interesting questions about the differing ways in which children

with ASD, young normal children and those with learning difficulties approach FB tasks.

### **Photographic cues.**

Photographic cues have also shown some success in facilitating FB performance. Charman and Lynggard's (1998) study with photographic cues is notable for finding no ASD-specific impairment on the standard FB task - although this appears to be because the two control groups (children with MLD and young TD children) performed particularly poorly, rather than because the group with ASD performed well. This study used a posting version of the Deceptive box Smarties task (adapted from Mitchell & Lacohee, 1991). When the child was asked 'What do you think is in the tube?' for the first time, they were invited to choose a photo of what they thought (i.e. Smarties) and post it into a toy postbox. The 'FB' question then took the following form, 'When you posted your picture, before we opened the tube, what did you think was in here?' This improved performance as compared to the standard task in all three groups (although not significantly in the group with learning difficulties). However, children with ASD still only performed at chance level, whilst over 70% of the two control groups passed the posting task, which was significantly better than chance. It is not demonstrated whether this facilitation shows any degree of transfer, and the usefulness of the cue appears to be very task-specific. Bowler and Briskman (2000) found that whilst the posting paradigm described above improved performance on the Smarties task in children with ASD and young normal children, the same technique had no effect on performance on an unexpected transfer test, even among children who had done a posting version of the Smarties task beforehand. This may reflect the relative difficulty of remembering someone else's belief, as opposed to your own (as in the Smarties task), or the greater complexity of the Sally Anne task.

This technique again removes some of the demands of the standard FB task. Firstly, the version used is really a 'memory for FB' task rather than a conventional FB task. In addition, reframing the FB question as 'When you posted your picture, before we opened the tube, what did you think was in there?' may have serious implications for the way the child understands the task. This question is substantially more elaborate than the standard 'Before I opened the tube, what did you think was inside?' In children with some degree of communication difficulties, their understanding of these questions may well not be complete, and they may be responding to the 'picture'



element of the question rather than to the 'thinking' element – in other words, it is possible that at least some of them are answering the question 'What was on the picture you posted?' Even if they do understand the 'think' element of the question, they do not necessarily have to consider mental states to reach the conclusion that the photograph is the right answer.

Another alternative is that this task removes some of the executive demands of the task – or at least, reverses the answer that a child with executive problems would give. Posting the picture may make the initial answer much more salient in the child's mind – and therefore whilst previously the salience of reality was greater than that of a prior mental state, now, the salience of a posted picture is greater than reality, particularly if the question specifically refers to that picture. However, the fact that Bowler and Briskman were unable to replicate these results using an unexpected transfer tasks seems to imply that it may be crucial that the task is one of memory for own FB rather than one of considering other's false beliefs, or that it eliminates some task-specific source of difficulty.

### **Training programmes designed to improve FB performance**

The various different training techniques and their efficacies are summarized in Table 6.1. Studies included are all those that have actively tried to train children or adults to pass FB tasks over a series of sessions (i.e. those that introduced modifications in the course of a single session are not included), and that tested the participants after the training on a separate task, rather than simply assessing progress during training. The different methods are then described, and their relative success discussed.

***Table 6.1. ToM training studies***

<b>Authors</b>	<b>Participants</b>	<b>Type of training</b>	<b>Results</b>
Flavell, Everett, Croft & Flavell (1981)	51 preschoolers (3-year-olds)	Individual. Visual perspective taking training.	No sig. improvement in performance on visual perspective taking tasks.
Taylor & Hort (1990)	49 preschoolers CA: 2:7 – 5:1	Individual. Played games designed to teach understanding of the terms 'looks like' and 'really and truly'.	No sig. improvement vs. pre-test or control group.
Ozonoff & Miller (1995)	9 adolescents with ASD (5 trained, 4 controls) CA 11:0 – 16:2, VIQ 66 – 104	Group. ToM and perspective taking skills. 14 sessions over 4 ½ months.	Trend towards sig. improvement on ToM composite. No improvement on parental social skills ratings.

Authors	Participants	Type of training	Results
Appleton & Reddy (1996)	46 3-year-olds (23 trained, 23 controls)	Group. Discussion of videos depicting UT FB scenarios. Controls: story reading sessions.	Trained group improved significantly on UT tasks, generalised to DB. Retained 2-3 weeks later.
Hadwin, Baron-Cohen, Howlin & Hill (1996)	30 children with ASD, in three groups of 10 (trained on emotion, belief or pretend play.) CA 4:0-13:0, VMA 5:0-11:0	Individual. Belief training: Principle-based training on 5 levels intended to mirror normal development.	Belief group: sig. improvement in number of levels passed at post test and 2 month follow-up. No generalisation between domains (e.g.. belief training did not improve emotion perception or pretend play).
Slaughter & Gopnik (1996)	Study 1: TD 3-year-olds in 3 groups of 11 (trained on belief, desires/perceptions and number conservation). CA 3:1 – 4:4. VMA not given	Individual. Belief: Feedback on A-R FB tasks. Desire/perception: Feedback on perspective taking and desire tasks. Two sessions over two weeks.	Sig. improvement on both self and other DB FB tests in both belief and desire/perception groups. No effect of control (number conservation) training on ToM tasks.
Slaughter & Gopnik (1996)	Study 2: Normal 3-year-olds in 3 groups of 13 (trained on belief, desires/perceptions and a language control task [controls]). CA = 3:3-4:5. VMA not given.	Same at study 1.	Sig. improvement in both belief and desire/perception groups on a range of FB tasks included self/other DB FB and appearance reality. Transfer seen to other tasks. No effect of control training on ToM tasks.
Steeremen, Jackson, Pelzer & Muris (1996)	Eight socially anxious and socially aggressive children, 3 control groups. CA 4:0 – 9:0 IQ of trained group 88-112.	Group. Stories, games and roleplays with ToM as a theme. Sessions over 21 weeks. One control group had non-ToM social-skills training programme.	Sig. improvement on FB tasks, emotion recognition and social skills rating. Not seen in control groups.
Swettenham (1996)	8 children with ASD, CA 5:6 – 5:10, VMA 3:1 – 4:2 8 children with DS CA: 5:9 – 15:6 VMA: 3:4 – 4:6 8 normal 3-year-olds CA: 3:2 – 4:4, VMA: 3:3-4:1	Individual. Feedback from computer of multiple trials of an unexpected transfer task. Two sessions a day over four days.	ASD: sig. improvement on close tasks only. Down/Normal: sig. improvement on close tasks and more distant (deceptive box). Results same at 3 month follow-up.
Swettenham, Baron-Cohen, Gomez & Walsh (1996)	8 children with ASD, CA: 8:9-14:4, VMA: 5:0 -:6:0	Individual. 'Photo in the head' strategy taught with a mannequin head and Polaroid camera. Training over 5 consecutive days.	7/8 passed UT task at follow-up (0/8 at pre-test). Some transfer to DB (3/8 passing) and S-K (6/8 passing).

Authors	Participants	Type of training	Results
McGregor, Whiten, & Blackburn (1998a)	16 individuals with ASD, (8 trained, 8 controls) CA: 8:6 – 3:9 VMA: 2:2 – 6:3 32 normal 3-year-olds (16 trained, 16 controls) CA: 3:0-3:8 VMA: 2:6 – 5:4	Individual. Intentions highlighted in a FB task and then a 'picture as thought' strategy was taught with dolls. Used only two scenarios repeatedly. Individual sessions.	ASD: Sig. Improvement on UT task (0/16 => 10/16), some transfer to DB task. 3-year-olds: At ceiling at UT task, some transfer to DB and to naturalistic FB.
MacGregor, Whiten & Blackburn (1998b)	5 adults with ASD CA: 22-39 VMA: 4:5 – 10:9 5 children with ASD CA: 9:6 – 17:3 VMA: 4:0 – 8:3	Individual. 'Picture as thought' strategy taught in the context of videoed FB tasks. Participants had already learned the strategy with dolls in earlier study (described above).	3/5 and 4/5 children passed at least 2/3 of the videoed FB tasks at post-test. 3/5 adults and 2/5 children had passed 1/2 of pre-tests.
Slaughter (1998)	30 preschoolers in 3 groups, trained in FB, false photo and false drawing, or number conservation. CA 3:3 – 4:1 VMA not reported.	Individual. Feedback on either FB scenarios, false photo and false drawing tasks or number conservation depending on training group. Two sessions over 2-3 weeks.	FB training sig. improved performance on FB and visual perspective taking. False photo training improved performance on false picture. No interaction between two. No effect of control training. (number conservation).
Ashcroft, Jarvis & Roberts (1999)	16 adults with learning difficulties CA: 28-68 VIQ not reported for all.	Group. Training using 'pictures in the head', acting FB scenarios with real people. Nine weekly one-hour sessions.	Sig. improvement on UT FB and transfer task
Knoll & Charman (2000)	Study 1: 22 preschoolers 11 trained 11 controls CA 2:11 – 4:2 VMA not reported.	Group. FB scenarios with feedback and discussion. 3 sessions over 2 weeks. Control: reading stories.	Sig. improvement on UT FB tasks only, no generalisation to DB. No improvement in controls.
	Study 2: 44 preschoolers. 15 FB trained 13 trained on visual perspective taking 16 controls CA 2:1 – 4:6	Group. FB training: same as above. Visual perspective taking training: scenarios and hiding games requiring the child to take differing perspectives, feedback and discussion as in the FB training. Control training: as above.	FB trained group improved on UT FB tasks only. Visual perspective taking group improved on visual perspective tasks only. No transfer. No improvement in controls.
Wellman, Baron-Cohen, Caswell et al. (2001)	Study 1: 7 male CWA. CA: 8-18 VMA 4:0 – 6:6	Individual. Six stage programme based on thought bubbles, using cardboard cutouts and story boards. Up to five individual sessions on consecutive days. Final stage included repeated FB tasks with feedback.	Improvement in UT type task (2/7 passed pre-test, 6/7 at post test). No transfer to DB task.

Authors	Participants	Type of training	Results
Wellman, Baron-Cohen, Caswell et al. (2001)	Study 2: 10 CWA. CA: 5-17 VMA 4:0 - 8:0	Individual. Training based on thought-bubbles (see study 1) over 3-8 sessions.	Sig. improvement over ToM aggregate. Fairly good transfer to DB task and S-K. Same tasks used at pre and post-test.

Abbs: ASD = Autistic spectrum disorder. FB = False Belief, ToM = Theory of Mind. UT = Unexpected Transfer FB task. DB = Deceptive Box FB task. S-K = Seeing Leads to Knowing task, A-R = Appearance Reality task.

### **Feedback on FB tasks.**

Straightforward feedback has had limited success with children with ASD, but appears to have been more useful with children with Down Syndrome or young TD children. Swettenham (1996) found that feedback from a computer on an unexpected transfer scenario enabled children with ASD to pass that task on the computer or the same task with dolls, but that they did not improve on a deceptive box task, unlike the TD children or those with Down Syndrome. There is additional evidence that feedback may be successful with young normal children. Slaughter and Gopnik (1996) describe a study in which three-year-old children were given feedback on appearance-reality type tasks. Two weeks after the training session, 54% of the group who were given feedback (referring to the beliefs of the characters) passed a FB task, whilst only 14% of the control group did.

### **Photographs or pictures in the head.**

More elaborate intervention schemes have aimed to teach children with ASD strategies to enable them to pass FB tasks. Photos or pictures in the head have been used as an analogy for thoughts in several studies. Swettenham et al. (1996) trained eight children with ASD on a programme that involved teaching them that people have photo-like thoughts in their heads. None of the children in this study managed to learn the rules for linking the photos with a mental state, although 7 of the eight children learnt rules linking the photos with the character's actions. Post-training, 7/8 of the children passed the 'Sally-Anne' FB scenario (from a baseline of all of them failing), 3/8 passed the Smarties FB test and 6/8 passed a 'Seeing leads to Knowing' task. The number who passed an appearance-reality task remained unchanged at 3/8. These results indicate a certain level of generalisation from training of this type, although since the children were not tested again at a later date we cannot tell whether they retained their understanding for any substantial length of time, nor whether the children are able to use the strategy outside the experimental situation.

McGregor et al. (1998a) also used a training programme that involved learning that people have thoughts in their heads like pictures. In a follow-up to that study, McGregor, Whiten and Blackburn (1998) trained 10 individuals with ASD on scenarios using real people. They used videoed versions of the FB tasks. The participants were initially asked to choose the photograph that showed the person's thoughts, then were helped with verbal prompts, and finally were shown a version with no prompts at all. At post-test, three (novel) videoed FB tasks were presented, and most of the adults (3/5) and children (4/5) passed two or more of these tasks. The authors interpret this as evidence that people with ASD can be helped to apply the 'thoughts in the head' technique to a real-life setting. However, there are several problems with this study. One is that only two FB tasks were administered at pre-test, and half of the participants passed one of these tasks. It is impossible to tell whether these participants would have passed or failed a majority of FB tests had they been given another task. If the five participants who passed one test at pre-test would in fact have passed 2/3 tasks had they been given the chance, only 2/10 participants would have gone from failing the majority of tasks administered to passing. In addition, although the authors claimed to have refined the technique in order to help people with ASD carry over their learning to real life scenarios, they made no attempt to assess this beyond the videoed FB tasks that closely resembled the scenarios used in the training. However, their findings do show that people with ASD can learn to apply a 'picture as thought' technique to real people, even though it is not possible to insert the pictures into real people's heads.

Other studies have used thought bubbles to highlight the thoughts of the characters involved in the FB scenario. Wellman et al. (2001) trained two separate groups of children with ASD on a programme that took place over between three and eight sessions of about 30 minutes each. Again they found improvement in performance on the unexpected transfer FB scenario, and some degree of transfer to other tests. Evidence from other studies (e.g. Parsons & Mitchell, 1999) indicates that children with ASD can learn to use thought bubbles as props to help them represent mental states.

### **Rule or principle based training.**

Rule-based strategies have also had some success. Hadwin et al. (1996) used a five stage programme designed to mirror the developing understanding of mind in

normal children. These five levels were simple perspective taking, complex perspective taking, seeing leads to knowing, predicting action on the basis of belief, and finally FB. Tasks assessed understanding at each of these levels, and the children were only training in the areas where they failed the task at the pre-test. The training programme took half an hour a day for up to eight days. The children were explicitly taught the principles underlying each stage, with the aim of improving the potential for generalisation. At the end of the training there was a significant improvement in the number of levels passed, and this was still significant, although somewhat attenuated, at a two month follow-up. However, it is not clear how many of the children were actually passing FB tasks, and how much of the improvement was due to improvement on simpler tasks such as a 'Seeing leads to Knowing' task. Hadwin et al. also trained children on emotion recognition and pretend play. Although the children in the Emotion group improved on the trained tasks, there was no indirect transfer of learning between any of the training programmes.

#### **Theory of Mind focused social skills training.**

Ozonoff and Miller (1995) used another rule-based approach within the context of a social skills training programme. The participants acted out role-plays that illustrated principles such as perception influencing knowledge and second order perspective taking. The programme also covered other areas such as conversational skills, and included a range of opportunities for the participants to practice their skills. ToM measures were used before and after the intervention. The intervention group improved substantially on the ToM tasks. Somewhat surprisingly, given the highly practical nature of the training, there was no improvement on parent and teacher rated social competence in the intervention group. This may reflect the insensitivity of the measure used to mentalising ability as opposed to social skills - items included inviting peers to the home, joining in group activities and initiating conversations, all of which do not necessarily require ToM skills, and may not be improved as a result of an improvement in mentalising ability.

#### **Theory of Mind training studies with non-autistic groups.**

A few training studies have concentrated on non-autistic groups who demonstrate difficulties on FB tasks. A range of ToM interventions appear to be remarkably effective in these groups – Steereman et al (1996) found a significant

improvement of ToM training on real life behaviour in socially anxious and aggressive children, and other studies have found that those with learning difficulties show a greater degree of generalisation to other ToM tasks than do children with ASD (Swettenham, 1996). Given that ToM difficulties are generally not thought to be fundamental in children with these types of difficulties, it is puzzling that a ToM intervention should have such wide reaching effect, particularly when the non-ToM social skills group have shown much less impressive progress (Steerneman et al., 1996). These two studies may indicate that difficulties in ToM may be more widespread than was initially thought, and whilst it may be the case that ToM is a fundamental deficit in ASD, difficulties may well also be present in other groups. Alternatively, it may be that individuals with learning difficulties fail these tests for task-specific reasons, which the training programmes enable them to overcome.

## **Summary**

A range of techniques has been used in an effort to improve performance on FB tasks. These include straightforward feedback (Swettenham, 1996), rule-based training (Hadwin et al., 1996), photos in the head or thought bubbles (Swettenham et al., 1996; Wellman, Baron-Cohen et al., 2001) and social skills training (Ozonoff & Miller, 1995). Whilst most of the studies can show improvement in a task similar to that used to train the children, fewer are able to demonstrate any degree of transfer to other tasks requiring a ToM. A straightforward feedback paradigm does not appear to show any degree of transfer to other types of FB task in children with ASD (although it does with young TD children and children with learning difficulties; (Swettenham, 1996). Other cued techniques are hard to evaluate since a range of tasks are not included. The strategy based training studies (Hadwin et al., 1996; McGregor et al., 1998a; Swettenham et al., 1996; Wellman, Baron-Cohen et al., 2001) discussed here have demonstrated fairly consistent success in improving the ability of children with ASD and those with learning difficulties to pass FB tests. The improvement in the strategy based training studies generally includes some degree of transfer, and the improvement has been shown in at least one case to be maintained over several months (Hadwin et al., 1996). Out of the three main strategies taught: pictures in the head; thought bubbles; and a rule-based strategy; there is no clear evidence that any one is significantly better than the others. However, there has been little attempt to

assess whether this improvement generalises to mentalising skills in every day life, and in many studies there is no assessment of whether the improvement is maintained beyond the end of the programme. Training programmes have also tended to focus almost exclusively on FB scenarios, played out by dolls or pictures, and have not (with the exception of Ozonoff et al; 1995), and some effort by McGregor et al; 1998b) related these scenarios to everyday life and real people. Additionally, these strategy based training studies have tended to concentrate on children with ASD alone, MLD alone, or in one case children with ASD contrasted with young TD children. This means that the opportunities for examining the differences between how children with ASD and those with learning difficulties respond to this type of strategy have not been exploited. Whilst comparisons with young TD children may be useful, it is debatable whether they provide a fair comparison group for children with ASD. For a start, they will mostly have no difficulties in learning, whilst most children with ASD have some degree of learning disability. This may have implications for how they respond to any training programme, as well as for how quickly they grasp the concepts involved. In addition, there is the difficulty that whilst very young normal children will not be able to grasp the requirements of the training programme, older children are likely to be on the point of passing FB tasks anyway. They may therefore be particularly receptive to instruction about ToM and false beliefs. Whilst children with learning difficulties who fail FB tasks are also unlike children with ASD in many ways, they can at least be matched on verbal ability with children with ASD, and are more likely to share difficulties in acquiring general concepts – a factor that may well be of crucial importance when participating in a training programme.

Training studies with individuals with learning difficulties or social problems other than ASD (Ashcroft et al., 1999; Steerneman et al., 1996; Swettenham et al., 1996) suggest that they may be highly receptive to ToM based interventions, and there is growing evidence that many children with learning difficulties do fail ToM tasks. Comparing these groups may give us more insight into the specific problems that children with ASD have, as opposed to more general problems common to all those with a learning disability. They may also give us insight into why some children with learning difficulties fail FB tasks – is it simply due to task demands such as memory, or do they genuinely have difficulties with mentalising, as demonstrated through understanding FB?



## **EF intervention studies**

In contrast to the ToM training studies, very little has been done in training executive functioning, and only one study has been done with children with ASD. Those that exist have concentrated mostly on children and adolescents with ADHD. Individuals with ADHD often demonstrate difficulties in sustaining attention and inhibitory control, although the exact nature of their deficits is not clear (see Sergeant et al., 2002 for a review). Interventions have been based around building up the children's visual and auditory attention skills. The EF training studies discussed are selected for their relevance to, and potential use with, children with ASD.

The only study to date to attempt to train any form of EF in children with ASD used the Tower of London task (Shimmon & Lewis, 2001). Participants in this study included children with ASD (12 trained, 12 controls), children with learning difficulties (9 trained, 9 controls) and TD preschoolers (11 trained, 10 controls). This study used a very brief training protocol, in which the children were trained on a single day by extra 'warm-up' tasks inserted before each set of tests, in which the participants were reminded of the rules and were asked questions such as 'which disc is best to move now?' or 'can you think of where you might move the yellow one now?' when they hesitated. The children were also given feedback at the end of each trial as to how they had performed. The control groups did the same task without the 'warm-up' trials and with no feedback. The post-test consisted of the Tower of Hanoi and a Card Sort task (a modified version of the DCCS). The children with ASD and TD children showed an effect of training on their performance on the Tower of Hanoi: 6/12 of the trained ASD group passing the 7-move version, compared to none of the control group. The effect on the TD children was smaller, with 6/11 of the trained group passing in contrast to 4/10 of controls. There was no significant effect of training in the MLD group. There was no generalisation to the card sort task, but this may have been because both the MLD and the TD groups were at or near ceiling on this task. A FB task was also administered before and after training (an deceptive box version before, an unexpected transfer version afterwards). There was no transfer of learning to the FB post-test. This study suffers from several limitations, most importantly the lack of a proper pre-testing of the task on which the children were trained. The pre-test in this study consisted only of a test of inhibition (the Luria Hand Game) and an deceptive box FB task, whilst the post-test consisted of the

Tower of Hanoi task and a card sort. The lack of a pre-test on the Tower of Hanoi makes it impossible to conclude that the trained group would not have performed better than controls at the Tower of Hanoi tasks *before* the training. In addition, the Card Sort and unexpected transfer FB tasks were not administered before training, meaning that we have to assume that the controls and trained groups would have performed at equivalent levels before training. The 'training' in this study mostly consisted of reminding children of the rules of the task and asking them questions that encouraged them to attend to the task. Whilst these are without doubt methods of improving task performance, it is debatable whether it actually improves their executive functioning, or whether it merely improved their attentiveness to the task. Finally, the scope of the study was very short term, with the children usually doing pre-tests and training on one day, and the post-tests the next day.

An improvement in performance on EF tests has been demonstrated in two studies with children with ADHD. Semrud-Clikeman, Nielsen, Clinton et al (1999) used visual and auditory attention tasks that gradually increased in difficulty over 32 hour-long group sessions. This was a very long term and intensive intervention project, with the children meeting twice a week in groups of 4 or 5 (with 2 or 3 adults) for eighteen weeks. The training consisted of the children doing the tasks repeatedly (whilst being timed) until they were 100% accurate on two consecutive occasions. Once they reached this level they moved onto the next, more advanced, task. They were not taught explicit strategies for improving their performance, but the tasks were discussed within each group and guidance was offered as to possible strategies. After doing the tasks, the children would verbalise their strategy and evaluate it based on their performance. Post-training, the group who had received the intervention had significantly improved on tests of visual attention and auditory attention when compared to a control group who also had ADHD. Indeed the intervention group's performance post-test was comparable to a group of controls who did not have attentional problems (and who did not receive any intervention). No efforts were made to assess the impact of the training on their everyday behavioural and attentional problems.

A second study, by Kerns, Eso, and Thomson (1999) also aimed to improve performance on attentional measures in children with ADHD. In this case the children were seen individually twice weekly, over eight weeks. The intervention consisted of training on sustained, selective, alternating and divided attention, and included both

auditory and visual activities. The visual materials were sets of cards showing pictures of people divided into 'families' on them. The children were asked to sort the cards in different ways, for example, by gender, by hair colour, or by family. The tasks became more difficult as the children progressed, and factors such as speed or distracting background noises were introduced. The children were not taught any specific strategies for improving their performance. A control group also met the experimenter for sessions, but they engaged in a range of computer-based games. Post-training, both groups improved significantly on a number of measures, but the intervention group had significantly larger gains on tasks requiring attentional flexibility (such as the Day-Night Stroop test), tests of selective attention, the Mazes subtest of the WISC-III, and a maths worksheet. There were no significant improvements on a symptom severity questionnaire completed by the parents and teachers of participants (the Attention Deficit Disorder Evaluation Scale, McCarney, 1989), although there was a trend towards improvement in the teacher's report of inattention and impulsivity for the intervention group.

The latter two studies discussed here were carried out with children with ADHD, and as such are not immediately appropriate for children with ASD. However, they do tell us that training children in EF is feasible, and give some suggestions as to possible techniques. Children with ASD appear to have difficulties in set shifting and flexibility (Pennington & Ozonoff, 1996), rather than in sustaining attention. The techniques used are also problematic - the training in the Semrud-Clikeman et al. study was extremely long term, and demands a very high level of motivation on the part of the participants, whilst the Kerns et al. study does not give any explicit instruction as to appropriate strategies to use. Given children with ASD's difficulties in perseveration and generativity, it is probable that under these circumstances they would find it very hard to devise strategies of their own. The Shimmom and Lewis study provides some evidence that children with ASD can be helped to improve their performance on tests of executive functioning, even with fairly minimal intervention. Certain aspects of the Kerns et al. study are potentially useful, and could be applied to training in flexibility and set shifting – for example having groups of cards which can be sorted by various dimensions. Training the children in shifting between these domains may help to improve their flexibility. It is also notable that the children in this group did show improvement on a test of

flexibility (the Day-Night Stroop task), unlike the children in the Shimmon and Lewis study.

## **Summary**

In conclusion, a number of studies have attempted to train children in theory of mind. Most of the children included have ASD, but some studies have trained other groups, including young TD children, children with DS, adults with MLD and socially anxious and aggressive children. Varying levels of success have ensued. All of the published studies show some improvement on FB tasks similar to that on which the children were trained, but there is less evidence for any degree of transfer to other tasks. Giving repeated feedback on FB tasks, in particular, showed no transfer of learning in children with ASD. In contrast, young normally-developing children and children with DS appeared to show a greater level of transfer of learning to other tasks. The strategy-based training studies were more successful, although neither the longer term retention of the learning, nor evidence for improvement in every day life social understanding has been well assessed. Many of the training programmes discussed also suffer from a lack of real life scenarios used in the training, focusing instead solely on dolls or mannequins. A lack of adequate controls may also be a problem; it seems unlikely that young TD children are a suitable control group for children with ASD.

The field of EF training is far less developed, and it is clear that a new programme would have to be developed to train children with ASD in flexibility and set shifting. However, the available studies with children with ADHD suggest some possible methods for doing so.

## **CHAPTER 7**

### **STUDY 3: A TRAINING STUDY OF THEORY OF MIND AND EXECUTIVE FUNCTION**

#### **Introduction**

As discussed in chapters 4 and 5, the relationship between EF and ToM has been the object of much recent research and debate. There are clear correlations between ToM and EF in young TD children, and children with ASD demonstrate difficulties in both ToM and EF, although these difficulties do not necessarily appear to be related, at least amongst those who fail FB tasks (Study 2). However, in ASD there is some argument over whether either EF or ToM is the fundamental deficit (Baron-Cohen, 1989; Russell, 1997b), and whether the difficulties seen in one area can account for the difficulties seen in the other. This study aims to investigate this relationship by training children with ASD in either ToM or EF (particularly set shifting and inhibition), and thus to look for interactions between the domains. In order to pursue the general theme of deficit specificity, children with ASD were compared with children with MLD. The hypothesis was that children with MLD do not have fundamental deficits in ToM, and therefore that they might benefit more from the training and show a wider degree of generalisation than children with ASD. Most training studies to date have failed to assess whether any gains are retained over time, and whether these gains are reflected in the child's behaviour in everyday life. This study therefore set out to train children in both ToM and EF, to assess the impact on experimental tasks both immediately and at follow-up, and to assess any impact on everyday behaviour, as rated by teachers.

The predictions for this study were as follows. Firstly, both children with ASD and those with MLD would be able to learn a simple strategy to enable them to pass FB tasks, but greater generalisation to other, non-trained, ToM tasks would be seen in the MLD group. Secondly, children with ASD and MLD would be able to learn a simple strategy to improve their performance on tests of cognitive flexibility such as card sorting tasks. Then there are four possible alternatives: (i) If ToM and EF are functionally independent, the training programmes will show no transfer from one area to the other. This would also be the case if they share underlying cognitive

factors, not addressed in the training programmes. (ii) If ToM is necessary for EF (either because of task-specific factors, an online need for a certain competence in one to facilitate the other, or developmentally), then the ToM training will lead to improvements in EF. (iii) If EF is necessary for ToM (because of task-specific factors, online or developmentally), then the EF training will lead to improvements in ToM. (iv) If ToM is necessary for EF, and EF is necessary for ToM, then the interactions will go both ways. Obviously this type of short duration training will not identify genuine long term developmental effects. However, within the 2-month time span of the follow-up, there may be sufficient time for more indirect (rather than immediate) effects of the training to become evident.

## **Methods**

### **Participants**

Participants were recruited from eight schools for children with special educational needs in the greater London area. They had all participated in Study 1 and Study 2, and were included on the basis of two criteria. First, that they had to have a VMA (as assessed by the BPVS II) of over 4 years 3 months. This was in order to ensure they had adequate language to understand the training programmes, and that they were over the age at which TD children would be expected to pass FB tasks. Secondly, they had to fail at least 2/3 FB tasks that involved predicting the contents of another's mind. These tasks are described in Study 1.

The children formed two groups. Twenty-one had non-autistic moderate learning difficulties (MLD), and 27 had autistic spectrum disorders (ASD). Twenty of the children in the latter group had a statement of special educational needs with ASD or autism as their primary diagnosis, one had a diagnosis of Asperger's Syndrome. The remaining 6 (who were from a single school and who were all being educated in classes for children with autistic spectrum disorders) were described on their records as having social and communication disorders. These children were assigned to the ASD group after completion of a checklist of symptoms based on DSM-IV after discussion with their teachers.

Children were randomly allocated to either one of two intervention conditions, the ToM training group, or the EF training group, or to a control group, who received no intervention. Children from the same school were distributed across groups, to

control for any school effects on the training. The CA, BPVS and TROG VMA and Raven's CPM (raw scores) are given in Table 7.1. The initial aim had been to have between 8-10 children in each experimental group. It proved very difficult to find appropriate children with MLD, since few such children with a VMA over 4 years consistently fail FB tasks, and so some group sizes are lower. Every effort was made to match for both school and VMA across these groups. However, in several cases children were allocated to the control group by default, by virtue of their having been absent from school over the weeks on which the training took place, and therefore the groups were not perfectly matched.

A one-way ANOVA showed no significant differences between the 3 experimental groups in each diagnostic group on any of the ability measures. As noted in Study 2, the ASD group as a whole differed from the MLD group in being significantly younger and having a significantly higher score on Raven's Matrices.

**Table 7.1 CA and ability of the six experimental group: mean (s.d).**

	ASD– ToM (n=10)	ASD– EF (n = 10)	ASD– Controls (n = 7)	MLD– ToM (n = 6)	MLD– EF (n = 7)	MLD– Controls (n = 8)
CA (years)	10.50 (3.02)	10.68 (2.68)	9.67 (1.73)	12.03 (1.77)	11.81 (2.20)	12.19 (1.32)
BPVS VMA (years)	7.23 (2.07)	6.57 (1.51)	5.44 (1.14)	7.46 (2.00)	7.08 (2.24)	6.90 (1.81)
TROG VMA (years)	5.00 (0.61)	5.35 (1.41)	4.49 (0.45)	5.83 (1.83)	5.57 (1.62)	5.19 (0.51)
Raven's CPM (raw scores)	22.90 (7.23)	24.60 (8.07)	20.57 (5.97)	16.00 (5.48)	16.57 (3.10)	18.50 (5.68)

### **Measures**

Pre- and post training measures were administered prior to beginning the intervention and within 2 days of completing the intervention. Follow-up testing was completed between 6 and 12 weeks later (mean time 9 weeks 4 days, s.d. 2 weeks 4 days), according to the schools' schedule and holidays. There were no differences in the delay between the training or control groups within each diagnostic condition (ASD  $F(2,24) = 0.45$ ,  $p = .64$ , MLD  $F(2,18) = 0.91$ ,  $p = .42$ ). There was a difference between the ASD and MLD groups ( $F(1,46) = 15.03$ ,  $p < .01$ ), with the ASD group having a longer delay. This was due to a particularly long delay at one school over the summer holidays. The six children from this school were equally distributed over all three experimental groups. Table 7.2 gives an overview of the design of the project, whilst

table 7.3 gives a summary of the tasks used at each time point, and what they were designed to test. The tasks were not all used at each time point in minimise exposure to the tasks and thus to lessen the chances of the children learning from task repetition.

**Table 7.2 Overview of project design**

Screening	Pre-test	Training	Post-test	Follow-up
Up to a month before training.	Immediately before training	20 minutes per day for 4-10 days	1-2 days afterwards	6-12 weeks afterwards

**Table 7.3 Tasks used at each timepoint**

Pre-test and screening	Post-test	Follow-up
ToM: FB unexpected transfer FB deceptive box self and other Pe-hiding deception Seeing leads to Knowing Knowing/guessing self and other	ToM: FB unexpected transfer FB deceptive box self and other Penny-hiding deception	ToM: FB unexpected transfer FB deceptive box self and other Seeing leads to Knowing Knowing/guessing self and other
ToM control: False photograph	ToM control: False photograph	ToM control: False photograph
EF: Flexibility and set shifting: Modified Wisconsin Card Sort Task ('Card Sort') Set shifting: Trails task Inhibitory control: Go-NoGo	EF: Flexibility and set shifting: Card Sort Inhibitory control: Go-NoGo	EF: Flexibility and set shifting: Card Sort Set shifting: Trails task
Everyday behaviour: Teacher questionnaire		Everyday behaviour: Teacher questionnaire

### **Screening**

The screening procedures are described in the methods section for Study 1 (Chapter 3). An extra unexpected transfer task was administered to those children who failed either the Sally-David task or the Smarties (other) task (but not both). Children had to fail either 2/2 or 2/3 FB tasks to be included.

### **Pre-training testing**

Children were given a pre-test to establish baseline measures of ToM, EF and general ability. Tasks used are showed in table 7.3, and described in full in studies 1 and 2. The pre-training testing session, which lasted approximately one hour, was administered by a single researcher in a quiet room in the child's school. In the case of some of the younger or least able children (ASD N = 16, MLD N = 4) this was divided into two half hour sessions to maximize their concentration.



### **Theory of Mind and Executive Function questionnaire:**

Teachers were asked to complete a questionnaire about the children both before training began and at follow-up. This questionnaire was designed using 15 items from previous studies (e.g. Frith et al., 1994) that had been considered to require a ToM. These included telling white lies, taking things literally, recognising surprise or embarrassment and responding to indirect hints. Fifteen items were included that were designed to be indicative of executive functioning (Booth, Happé, Hughes, & Charlton, in prep) for example, having an ability to plan ahead, being able to do mental arithmetic and being able to follow verbally given lists of instructions. An additional five items were included which were social items that were thought not to necessitate a ToM, for example having appropriate table manners, and recognising happiness and sadness. These items went to form the 'general sociability' (G-S) scale. A scoring system similar to that used by Frith et al. (1994) was utilised: 2 points for behaviour 'definitely' shown, 1 point if 'sometimes' or 'rarely' shown (and for don't know answers) and 0 for behaviours shown 'not at all'. These scores were compiled to give 3 scales, two of which (the ToM and EF scales) ranged from 0-30, and one of which (G-S) ranged from 0-10. See Chapter 9 for alpha values for these scales, and for a more detailed account of the questionnaire distribution. See Appendix II for a list of all items in the questionnaire.

### **Training programmes**

There were two training programmes. Both consisted of 25-minute long individual sessions held on consecutive days. The training was to criterion, and so the children could take between 4 and 10 days to complete the programme. At the end of each day, the child and the researcher discussed what had been learnt, and wrote a sentence on a 'reminder card'. Each new session began with reading the sentence on the card, and discussing what had been learnt the previous day. Brief outlines of the training programmes are given below, and fuller descriptions for each training programme are shown in Appendix III and IV.

#### **ToM training**

This training programme was based on that used by Swettenham et al. (1996). However, the programme was modified in a number of ways, as follows. A doll was used instead of a mannequin's head. Illustrative stories about children in similar

situations to those used in the training that day were introduced at the end of most stages, and two perspectives were introduced in stage 3. The training was to criterion, and presentations were unlimited up to a maximum of ten 20-25 minute sessions. If a child failed to reach criterion on one day, the stage was repeated with different materials the next day until they passed it. The photos were referred to throughout as 'thought-pictures', and were linked explicitly to mental states throughout the training.

### Design

The five rules to be learnt at each stage of the training were as follows:

Stage 1: Introduction of camera analogy. 'When a person sees something, they have a thought in their head and we can say it's like a picture.'

Stage 2: 'Thought pictures can stay in a person's head even if they go away. They can use the thought picture to look for things.'

Stage 3: 'Different people can have different thought pictures in their heads'

Stage 4: 'Sometimes thought pictures can be out-of-date.'

Stage 5: 'A thought is like a thought-picture, only we can't see it.'

Each of the stages consisted of demonstrations and questions to check the child understood the principles involved. The child was allowed to move onto the next stage once they had answered 3 consecutive trials correctly. At the end of some of the stages illustrated stories were given about real children in scenarios that paralleled the concepts used at that stage. The children did not have to pass these to move onto the next stage, but their responses were recorded. The training was designed to last between 5 and 10 days. In practice, some children completed it in 4 days, as they completed Stages 1 and 2 in a single session. The longest duration was 8 days.

### Materials

Two dolls with plastic heads were adapted for the training. Slots were cut in the top of the dolls' heads so that Polaroid photos could be inserted. See Appendix III for a photograph of the dolls. Some Polaroid photos were prepared beforehand; others were taken during the course of the training. A Polaroid 'JoyCam' was used. Objects used during the training were a purple cylinder, an orange basket, two coloured boxes, an Elmo toy puppet, a toy Mr Happy, a toy crocodile, a toy elephant, a small teddy bear. See Appendix III for pictures of the materials.

The stories were illustrated with line drawings.

## **Executive Function training**

### **Design**

No appropriate EF training programmes for children with ASD were found in a literature search on the topic. Several programmes exist for children with ADHD, and whilst these programmes are generally not relevant for children with ASD, or are too long term for the scope of this project, elements from these studies were used, such as the use of cards which could be sorted by different dimensions (Kerns et al., 1999). The programme was designed to target flexibility and set shifting, aspects of EF that appear to be impaired in ASD. The EF training paralleled the ToM programme as closely as possible, having five stages on which the child was trained to criterion. Whilst the ToM programme used a ‘thoughts as pictures’ analogy, this programmes used a ‘brain as machine’ analogy, introducing the idea of using different ‘brain tools’ for different activities. The five rules to be learnt were as follows:

- Stage 1: People can do lots of things. Sometimes they need to change how they do things. They can do this by changing their brain tools.
- Stage 2. Sometimes if we have been doing something for a long time, it’s hard to change our brain tool.
- Stage 3. Some brain tools are easier to use than others. Stop-Change-Go sequence for changing brain tools.
- Stage 4. Sometimes we have to change our brain tools before we finish doing something.
- Stage 5. Sometimes we have to decide for ourselves what brain tools to use.

Each session consisted of a demonstration and practice, using cards. The child had to perform to criterion in order to move on to the next section. These criteria were set using the binomial distribution, and were calculated so that there was less than a 5% probability of the child achieving that result by chance alone. The concept of changing tools was introduced using a toy truck, which had a range of changeable fitments that could be used to pick up stones, sand etc. There were illustrated stories at the end of some stages, drawing parallels between the exercises and real life. A ‘reminder card’ was completed at the end of each session, and was read at the beginning of each new session to refresh the child’s memory of the previous exercises. A full description of the training is given in Appendix IV.

## **Materials**

A scale model yellow CAT truck with removable tools was used. A toy plastic car, some twigs and a collection of small stones were used with the truck. A laminated, line drawn cross section of a head, showing the brain, was used as the 'brain card'. 'Brain tools' were oval laminated pieces of card with descriptions and illustrations for those who couldn't read (e.g. shapes on the 'Shape' tool, numbers on the 'Number' one). Several sets of laminated cards were used. These varied along different and differing numbers of dimensions. The simplest ones varied along only two dimensions, e.g. yellow and red circles and rectangles. The more complicated ones used complex shapes and more dimensions, e.g. card colour, type of animal, number of animals present. See Appendix IV for pictures of the materials used.

## **Post-training testing and follow-up**

The post-training testing session was administered one or two days after completion of the training programme, and the follow-up tests were administered between 6 and 12 weeks later. An outline of the tasks used at each time point is given in table 7.3. In order not to repeat tasks too many times (and therefore to minimize practice effects), a reduced number of tasks were administered at this stage. Different materials were used for each version of the task administered. Tasks were presented in a single half hour session at each timepoint. Many of the task protocols are described in the results section of Study 2 (chapter 5).

## **Results**

Results will be reported under subheadings, reflecting the key research questions. The first issue to be considered is the response to the two types of training. The second issue concerns the indirect effects of training – whether intervention in ToM improved performance in EF or vice versa.

### **Effects of training**

#### **Baseline scores**

Table 7.4 gives baseline scores for the 6 experimental groups on the ToM total score (made up of a sum of all seven ToM tasks), the Eyes task and EF measures (Card Sort aggregate score, Trails Difference score, Go-NoGo A'). The Card Sort aggregate score was calculated from 5 variables; percentage perserverative errors, failure to maintain set, percentage conceptual level processing, number of trials to

complete first category and number of categories achieved. For further details of this aggregate see Study 2. For a break down of the Card Sort results variable by variable see Table 7.9.

**Table 7.4. Baseline scores on the ToM total score, the Eyes task and EF measures for the six experimental groups. Mean (s.d.)**

	ASD- ToM	ASD- EF	ASD – Control	MLD – ToM	MLD – EF	MLD – Control
ToM aggregate	2.40 (1.51)	2.10 (1.29)	2.29 (0.95)	3.33 (1.63)	3.00 (1.53)	4.13 (1.64)
Eyes task	3.90 (1.91)	5.00 (3.27)	3.86 (0.69)	4.50 (1.97)	5.57 (2.26)	5.13 (2.30)
Card Sort aggregate	4.60 (2.22)	5.80 (3.29)	3.86 (2.79)	4.17 (2.32)	4.57 (2.15)	4.00 (1.51)
Trails	78.26 (48.98)	45.88 (34.73)	36.65 (25.63)	89.32 (93.78)	63.43 (44.31)	63.85 (51.75)
Go-NoGo A' <sup>a</sup>	0.82 (0.09)	0.73 (0.12)	0.72 (0.19)	0.74 (0.12)	0.76 (0.10)	0.83 (0.12)

<sup>a</sup> Due to computer malfunction, Go-NoGo results were not available for 10 of the children in the ASD group. This affected 3 children in each of the training groups and 4 controls.

One-way ANOVAs were performed to compare the three experimental groups within each diagnostic category. No significant differences were found between groups (in the ASD group all p-values >.16, in the MLD group all p-values >.31). Comparisons of the separate Card Sort variables also showed no significant differences (all p-values >.24).

### **Effect of training on ToM performance**

Figure 7.1 shows the pre-test, post-test and follow-up scores on the ToM aggregate for the ASD and MLD groups separately. As slightly different tests were given at post-test and follow-up, the percentage of tests passed is used rather than raw scores. Table 7.5 gives these proportions and standard deviations.

Due to the small size of the groups and the non-parametric nature of the data, one tailed Wilcoxon's Signed Ranks tests were used to compare performance at the different time points. Significant improvements as compared to the pre-test are given in table 7.5. Between the post-test and the follow-up, significant improvements were made by the ASD–EF group ( $z = -2.12$ ,  $p < .05$ ) and the MLD–EF group ( $z = -2.05$ ,  $p < .05$ ).

**Table 7.5. Mean percentage of ToM tasks (s.d) passed at each stage for each experimental group.**

	ASD – ToM	ASD – EF	ASD – Controls	MLD – ToM	MLD – EF	MLD- Controls
Pre-test	34.29 (21.51)	30.00 (18.38)	32.65 (13.59)	47.62 (23.33)	42.86 (21.82)	58.93 (23.46)
Post-test	60.00** (24.15)	37.50 (29.46)	39.29 (28.35)	83.33* (20.41)	50.00 (28.87)	78.13* (20.86)
Follow-up	53.33* (28.11)	51.67** (19.95)	30.95 (15.00)	69.44* (30.58)	66.67* (30.43)	79.17* (24.80)

\*  $p < .05$  \*\*  $p < .01$  as compared to performance at pre-test.

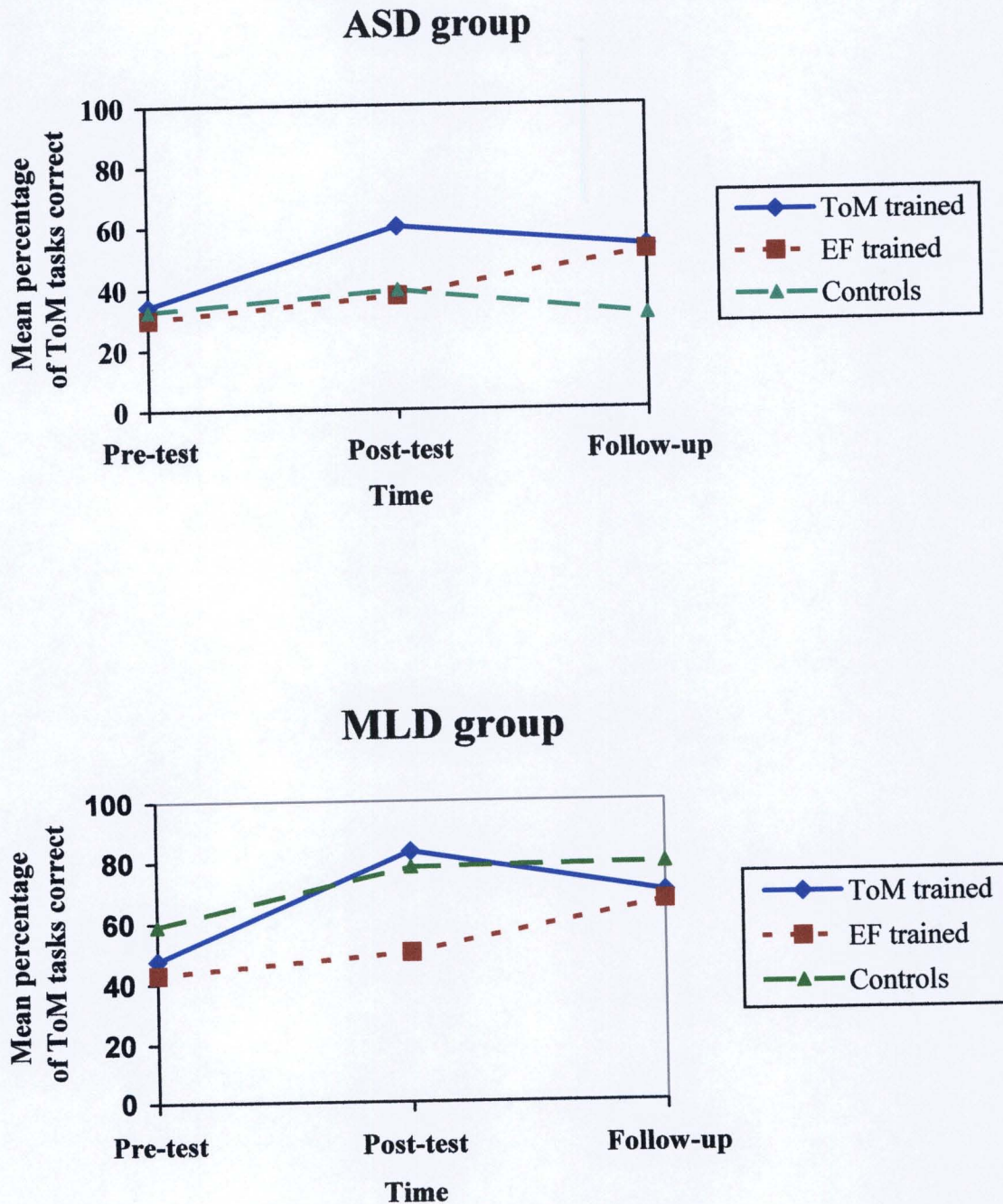
### **Generalisation to ToM tasks other than the trained task.**

Results on individual ToM tests were compared using one-tailed Wilcoxon's Signed Ranks tests in order to examine whether the improvements seen in the ToM group were due solely to improvement on the trained task. The unexpected transfer task used at post-test and at follow-up was a novel task, but one which had the same underlying structure as those used in training.

Between the pre-test and the post-test, the ASD-ToM group improved significantly on 2 out of the 4 tasks administered, the trained 'unexpected transfer' task ( $z = -2.65$ ,  $p < .01$ ) and the penny-hiding task ( $z = -1.89$ ,  $p < .05$ ). The EF group improved significantly on the penny-hiding task only ( $z = -2.24$ ,  $p = .01$ ). The ASD-Control group improved on the Deceptive box-other task ( $z = -1.73$ ,  $p < .05$ ).

In the MLD groups, the ToM group improved on 3 out of the 4 tasks administered, the trained task ( $z = -2.24$ ,  $p < .01$ ), the Deceptive box – other ( $z = -2.00$ ,  $p < .05$ ) and the penny-hiding ( $z = -1.73$ ,  $p < .05$ ). The EF group did not improve significantly on any individual tasks; the Controls improved on the unexpected transfer task ( $z = -2.45$ ,  $p < .01$ ) and the penny-hiding task ( $z = -1.89$ ,  $p < .05$ ).

**Figure 7.1: Graphs showing proportion of ToM tests passed at pre-test, post-test and follow-up in the experimental groups**



Between the pre-test and the follow-up, the ASD-ToM group improved significantly only on the trained task ( $z = -2.12, p < .05$ ). The ASD-EF group improved on Deceptive box-self ( $z = -1.63, p < .05$ ), Deceptive box-other ( $z = -1.73, p < .05$ ) and Seeing Leads to Knowing ( $z = -2.00, p < .05$ ), and they got worse on the Know/guess self ( $z = -1.73, p < .05$ ). On closer inspection of this task it was clear that

five members of this group had performed inconsistently on this task (i.e. passing the ignorant condition whilst failing the knowledgeable, or vice versa), and one individual had failed due to refusing to answer. The ASD-Controls improved on nothing.

The MLD-ToM group improved on the trained task ( $z = -2.00$ ,  $p < .05$ ), and on Deceptive box-other ( $z = -1.73$ ,  $p < .05$ ). The MLD-EF group improved on no individual tasks, but there was a trend towards an improvement on the Know-guess/self and Know-guess/other ( $z = -1.41$ ,  $p = .08$  in both cases). The MLD-Controls improved on the (un) trained task only ( $z = -2.24$ ,  $p = .01$ ). See Appendix V for tables of the number of children who passed each task at each timepoint. There was no improvement in any group at any timepoint in the false photograph task, but this may have been because performance at pre-test was so high in both groups (over 70% of each group passing).

The percentage of children in each group who improved on at least one ToM task (i.e. passed more ToM tasks at post-test or follow-up than they did at pre-test) is shown in Table 7.6. The number in brackets gives the percentage who improved on a ToM task other than the trained unexpected transfer task.

**Table 7.6 Percentage of children in each group who improved on the proportion of ToM tasks passed between pre-test and post-test or pre-test and follow-up (percentage improving on a task other than the unexpected transfer task).**

	ASD- ToM	ASD- EF	ASD- Control	MLD- ToM	MLD- EF	MLD- Control
Pre-test => post-test	80 (60)	50 (60)	43 (29)	100 (83)	57 (57)	88 (75)
Pre-test => Follow-up	60 (50)	50 (50)	14 (14)	50 (33)	43 (43)	63 (25)

It may seem counter-intuitive that a higher percentage of the ASD-EF group improved when the unexpected transfer task was removed from the analysis than when it was included. This is because the improvement was calculated by subtracting performance on the tasks at pre-test from that at post-test. A single individual in the ASD-EF group passed the unexpected transfer task at pre-test and failed it at post-test, but passed the Deceptive box -self test at post-test, having failed it at pre-test. When all the tests were included, therefore, he did not appear to have improved as he passed one test at pre-test and one test at follow-up. However, once the unexpected transfer task was removed, he appeared to have failed all tasks at the pre-test but have passed one at post-test, and was therefore counted as having improved.



### Eyes task

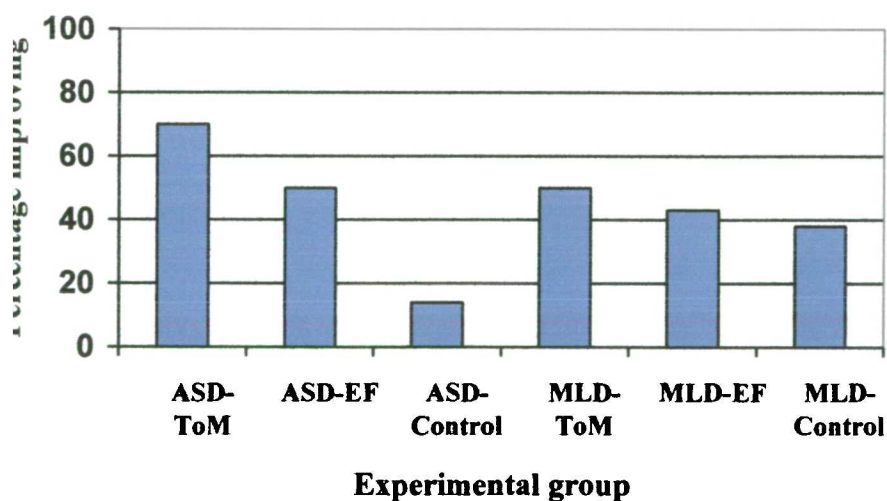
The Eyes task was not included in the ToM aggregate, as it has a continuous rather than categorical score, and did not relate to the other ToM tasks (see chapter 5). Table 7.7 gives the pre-test and follow-up scores for the Eyes task (it was not given at post-test).

**Table 7.7 Performance on the Eyes task at pre-test and follow-up: mean (s.d.)**

Eyes task (max = 14)	ASD- ToM	ASD- EF	ASD- Control	MLD- ToM	MLD- EF	MLD- Control
Pre-test	3.90 (1.91)	5.00 (3.27)	3.86 (0.69)	4.50 (1.98)	5.57 (2.64)	5.13 (2.30)
Follow-up	5.10 (1.60)	5.30 (2.21)	4.29 (1.38)	5.33 (2.25)	6.43 (2.15)	5.63 (2.20)

Wilcoxon signed rank tests showed no significant improvements in the group means. To look at improvements on a more individual level, the number of children who had improved on the task was calculated for each group. Figure 7.2 shows the percentage of children in each group who improved between pre-test and follow-up.

**Figure 7.2 Percentage of each group improving on the Eyes task between pre-test and follow-up**



Fisher's Exact test found a significant difference between the ASD-ToM group and the ASD-Controls only ( $p < .05$ ).

## **Effects of training on EF performance**

### **Card Sort**

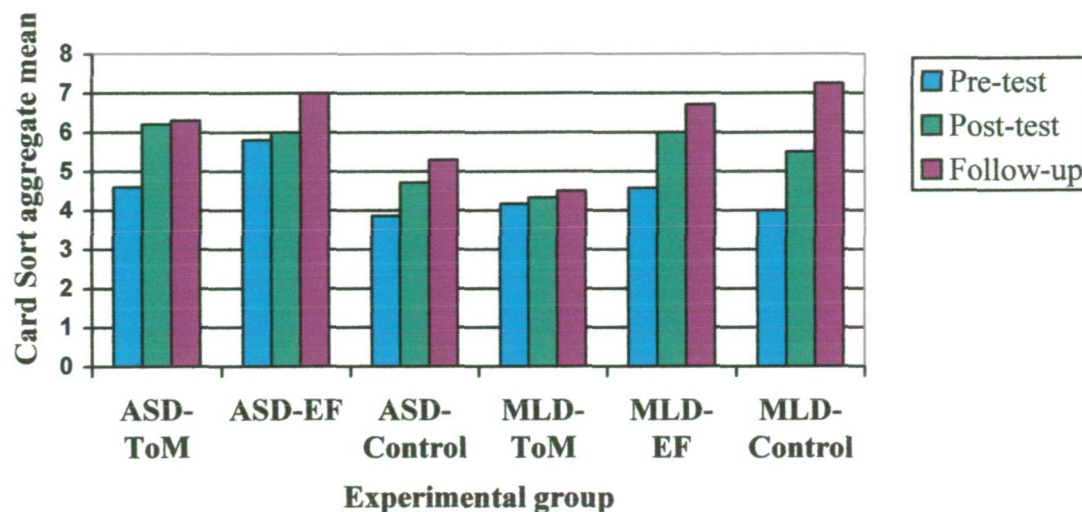
Table 7.8 gives the groups' scores on the various Card sort measures at pre-test, post-test and follow-up, and figure 7.3 shows the mean scores on the card sort aggregate at each time point.

***Table 7.8 Scores on Card Sort variables for the six experimental groups at pre-test, post-test and follow-up***

		ASD- ToM	ASD- EF	ASD- Control	MLD- ToM	MLD- EF	MLD- Control
Percentage perserverative errors (low = good performance)	Pre-test	41.39 (23.61)	31.32 (23.93)	43.49 (22.41)	30.22 (22.25)	25.33 (20.11)	25.62 (15.08)
	Post-test	25.19* (21.41)	26.72 (23.43)	32.90* (24.00)	34.67 (12.27)	17.58 (6.07)	22.13 (15.25)
	follow- up	26.43 (21.55)	20.01 (16.26)	33.57 (23.75)	32.81 (18.88)	19.21 (15.05)	13.18* (7.49)
Number of trials to first category (low = good performance)	Pre-test	11.25 (6.84)	9.67 (4.21)	6.33 (0.58)	8.80 (4.21)	9.40 (3.91)	11.50 (5.01)
	Post-test	9.44 (4.59)	8.50 (2.56)	7.00 (1.00)	8.50 (3.79)	14.33 (4.50)	12.50 (6.95)
	follow- up	8.30 (5.58)	9.67 (6.00)	8.00 (3.52)	7.50 (2.74)	8.14 (2.91)	7.00 (1.41)
No. of times failed to maintain set (low = good performance)	Pre-test	0.20 (0.42)	0.20 (0.42)	0.14 (0.38)	0.67 (0.82)	0.43 (0.79)	0.50 (0.53)
	Post-test	0.40 (0.52)	0.40 (0.52)	0.14 (0.38)	0.33 (0.52)	0.29 (0.49)	0.13 (0.35)
	follow- up	0.40 (0.70)	0.40 (0.70)	0.71 (0.95)	0.67 (0.82)	0.28 (0.49)	0.13 (0.35)
No. of categories achieved (high = good performance)	Pre-test	1.90 (1.29)	2.50 (1.58)	1.29 (1.38)	1.67 (1.21)	1.57 (1.51)	2.13 (1.25)
	Post-test	2.20 (1.39)	2.50 (1.43)	1.42 (1.51)	1.17 (1.17)	2.43 (0.98)	2.00 (0.76)
	follow- up	2.50 (1.18)	2.90 (1.19)	2.29 (1.38)	1.67 (0.82)	2.43 (1.27)	2.63 (1.19)
% conceptual level sorting (high = good performance)	Pre-test	25.25 (18.04)	36.05 (26.10)	16.33 (13.89)	30.47 (23.16)	31.09 (26.13)	28.78 (16.41)
	Post-test	43.64* (31.89)	40.91 (24.00)	20.65 (21.28)	20.25 (11.21)	38.36 (18.88)	27.05 (13.83)
	follow- up	39.44 (27.16)	46.56 (20.03)	34.25* (23.20)	22.14 (13.24)	44.69 (31.56)	42.01 (27.64)

\*p<.05 when compared to pre-test scores.

**Figure 7.3. Performance of each experimental group on the Card Sort aggregate at each time point**



To summarise the findings in table 7.9 between pre-test and the other time points:

**At post-test:** Wilcoxon Signed Rank tests found no improvement on any of the card sort variables in the ASD-EF group (all  $p$ -values  $> .20$ ). The ASD-ToM group improved between the pre-test and post-test on percentage perserverative errors ( $z = -1.96, p=.05$ ), as did the ASD-Controls ( $z = -2.37, p<.05$ ).

In the MLD group there were no significant improvements between pre-test and post-test.

**At follow-up:** In the ASD group, the ToM group improved on percentage conceptual level sorting ( $z = -2.09, p<.05$ ), as did the controls ( $z = -2.37, p<.05$ ). There were no improvements in the EF group.

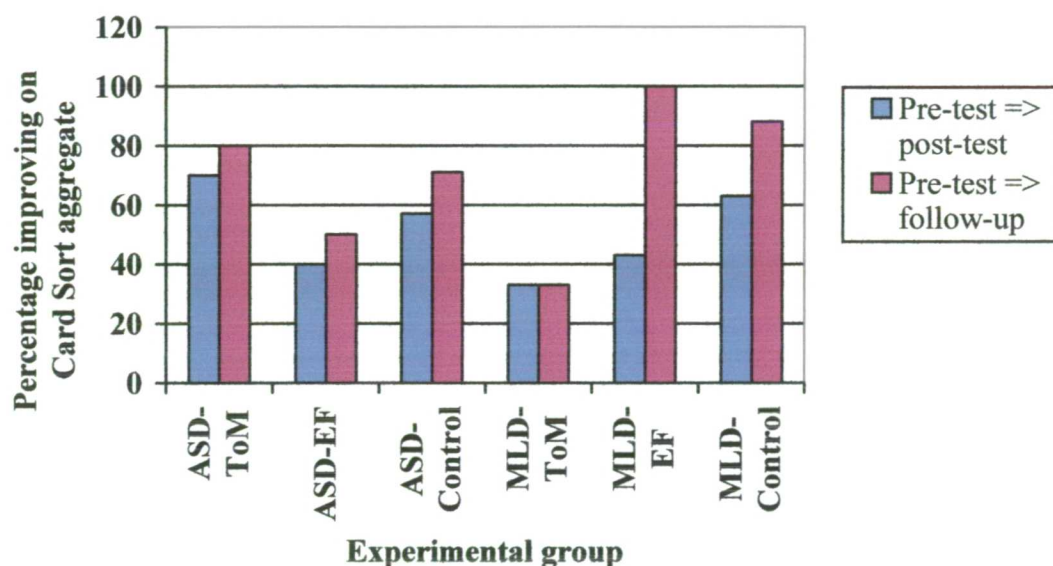
In the MLD group, the EF group had a trend towards improvement on the number of categories achieved ( $z = -1.86, p=.06$ ), and the Control group improved on percentage perserverative errors ( $z = -2.10, p <.05$ ) and had a trend towards improvement on all other categories (all  $p$ -values  $<.10$ ).

**Card Sort aggregate:** There was a trend towards an improvement in the ASD-ToM group between pre-test and post-test ( $z = -1.92, p=.06$ ) and a significant improvement at follow-up ( $z = -2.15, p<.05$ ). There was no improvement in the ASD-EF group and a trend towards improvement between pre-test and post-test in the ASD-Controls ( $z = -1.86, p=.06$ ).

In the MLD group, there was no improvement in ToM group, there was an trend towards improvement in the EF group between pre-test and post-test ( $z = -1.63$ ,  $p=.10$ ) and a significant improvement between pre-test and follow-up ( $z = -2.41$ ,  $p<.05$ ). These results were mirrored by the MLD–Controls (pre-test => post-test,  $z = -1.87$ ,  $p=.06$ , pre-test => follow-up  $z = -2.39$ ,  $p<.05$ ).

In order to look at differences on a more individual level, the percentage of children in each category who improved on the card sort aggregate between the time points was examined (figure 7.4).

**Figure 7.4 Percentage of each group improving on the Card Sort aggregate between pre-test, post-test and follow-up**



Fisher's exact test showed no difference between the groups.

### **Trails and Go-NoGo**

The Go-NoGo was given only at post- test and not at follow-up, and the Trails test was given only at follow-up. Table 7.9 gives the before and after training scores for both the Go-NoGo and the Trails differences score (made by subtracting the time for Trails A from the time for Trails B).

On the Trails task, Wilcoxon Signed Ranks tests showed a trend towards improvement in the ASD–ToM group ( $z = -1.68$ ,  $p = .09$ ), and the MLD-EF group got significantly worse ( $z = -2.02$ ,  $p <.05$ ).

A fairly large group of children were not included in this analysis because they had not completed the Trails task within the 300 second time limit at either the pre-test or the follow-up. In order to include them, an alternative score was created which

was of improvement between the time points. Children classed as having ‘got worse’ either if their time increased from pre-test to follow-up, or if they had completed the tasks at pre-test but failed to do so at follow-up. They were counted as staying the same if they completed the task in the same time to within five seconds, or failed to complete the task both times, and were counted as improving if they either got faster, or completed the task the second time having failed to do so at pre-test. Table 7.10 gives the percentages (rounded to the nearest unit) of children in each group who improved, stayed the same or who got worse. Fisher’s exact test found no significant differences between any of the experimental groups.

**Table 7.9 Performance on Go-NoGo and Trails at pre-test, post-test or follow-up**

		<b>ASD– ToM</b>	<b>ASD– EF</b>	<b>ASD– Controls</b>	<b>MLD– ToM</b>	<b>MLD– EF</b>	<b>MLD– Controls</b>
Trails difference score <sup>b</sup>	Pre-test	78.26 (48.98)	45.88 (34.73)	36.65 (25.63)	89.32 (93.78)	63.43 (44.30)	63.85 (51.75)
	follow- up	60.83 (46.05)	41.69 (34.54)	42.06 (42.87)	71.29 (61.13)	94.55 (61.09)	57.77 (34.53)
Go-NoGo A’ <sup>a</sup>	Pre-test	0.82 (0.09)	0.73 (0.12)	0.72 (0.19)	0.74 (0.12)	0.76 (0.09)	0.83 (0.12)
	Post- test	0.83 (0.13)	0.78 (0.19)	0.72 (0.12)	0.78 (0.16)	0.76 (0.13)	0.83 (0.13)

<sup>a</sup> Due to computer malfunction 10 children with ASD are not included in these scores. These include 3 children in each of the training groups and 4 children in the control group.

<sup>b</sup> Some children were unable to complete the Trails B within 300 seconds and so the test was abandoned. This affected 2 children in the ASD-ToM group at pre-test and one at follow-up, one child in the ASD – EF group and 2 at follow-up, 3 children in the ASD – Control group at pre-test and one at follow-up, 3 children in the MLD – ToM group at pre-test and one at follow-up, one child in the MLD-EF group at pre-test and none at follow-up, and one child in the MLD – Control group at pre-test only.

**Table 7.10 Percentage of children improving or not between pre-test and follow-up on the Trails task**

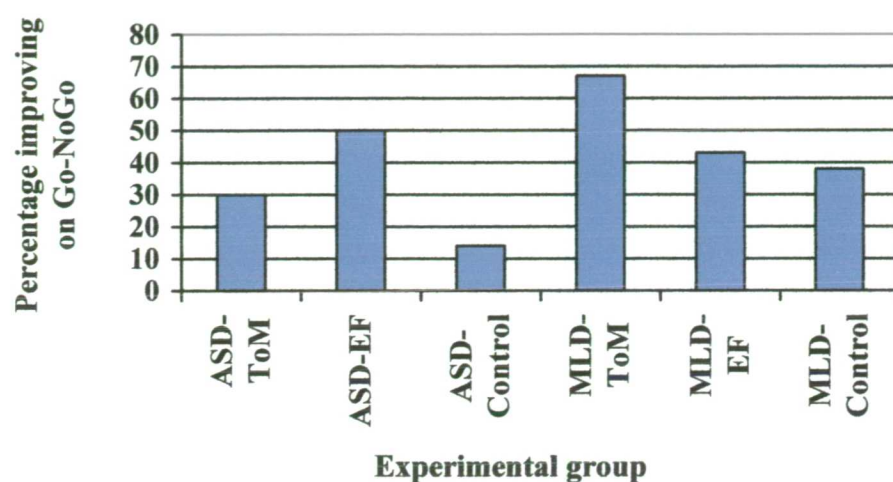
	<b>ASD– ToM</b>	<b>ASD– EF</b>	<b>ASD– Controls</b>	<b>MLD– ToM</b>	<b>MLD– EF</b>	<b>MLD– Controls</b>
% improving	10	40	14	17	71	38
% no difference	30	30	14	17	29	13
% got worse	60	30	71	67	0	50

### Go-NoGo

Wilcoxon’s signed ranks test found no significant differences between any groups on the Go-NoGo A’ score (all p-values > .24). Figure 7.5 shows the percentages of each group who improved on the Go-NoGo task. There were no significant differences between the experimental groups within each diagnostic category.



**Figure 7.5 Percentage in each group who improved on the Go-NoGo task between pre-test and post-test**



### **Summary of pre-test, post-test and follow-up results**

#### **Theory of Mind**

It appears that the ToM training worked in that the children in the ToM group all learnt how to pass the trained task, and there was some degree of transfer of learning to other tasks. Both the ASD- and MLD-ToM trained groups improved on the penny-hiding deception task at post-test, a task in which no training had been provided. The MLD-ToM group also improved on the Deceptive box-other task, both immediately and at follow-up. The ASD-ToM group only improved on the trained task at follow-up – this may have been due to the exclusion of the penny-hiding task at this time point, rather than due to reduced generalisation of learning across time. In general there was a slight drop off in the proportion of ToM tests passed by these groups between the immediate post-test and follow-up. There was some tentative evidence for a slight improvement on the Eyes task in the ASD-ToM group.

The EF trained groups, both ASD and MLD, also made significant improvement on the ToM tasks, but this was over a longer time frame - whilst they made no significant gains on the ToM aggregate between the pre-test and post-test, they had made significant gains by the follow-up as compared to the pre-test, and also made significant gains between the post-test and the follow-up. They thus appeared to follow a different trajectory to the ToM trained groups, making more gradual gains over time. Neither of the EF trained groups (ASD or MLD) made gains on the

unexpected transfer task in which the ToM group had been trained, instead most of their gains were on the other tasks – the Deceptive box–self in both groups, and the Deceptive box-other and Seeing Leads to Knowing in the ASD group. The ASD-EF group therefore showed a wider improvement on a range of ToM tasks than did the MLD-EF group.

The MLD controls appeared to make consistent improvements without any training. They started off with the highest mean score on the ToM aggregate (4/7) of any of the groups, and they made a significant improvement on the unexpected transfer task. Both the post-test and follow-up unexpected transfer tasks used dolls rather than the illustrated pictures of the pre-test, and it is possible that this group had found the illustrations a particularly difficult medium, thus masking their competence which was revealed in the other tasks. The ASD controls made small gains at post-test but had reverted back to their earlier levels by the follow-up.

### Executive Function

There was no real evidence that the EF training improved performance on any of the EF post-test or follow-up tasks. Any improvements seen were generally inconsistent across time and no different to those seen in the control groups. The ASD-ToM group improved on some of the card sort variables, but this was mirrored by the ASD-Control group, making it likely this was either spurious or due to task practice. The EF trained groups did, however, show improvement on the ToM tasks. This improvement may have been due to chance (although this is unlikely since it was statistically significant despite very small numbers), or to non-specific aspects of the training programmes, for example improved familiarity with the experimenter. In order to investigate whether in fact any specific – i.e. EF related - elements of the training programme might have lead to the improvement seen in ToM, the relationship between performance on the EF training programme and the improvement in performance on ToM tasks was examined.

### Executive Function training and Theory of Mind

As the children in the EF training group were trained, their responses were recorded. These responses were used to create two variables that gave a measure of their response to training. The first was ‘Attempts’; a sum of the number of attempts taken by a child to meet the criterion to pass each stage. There were five different assessment points, and so the scores could potentially range from 5 (passing on the

first attempt at each stage) upwards. In reality the overall mean was 6.9 (s.d. = 2.5), with a range of 5 – 13. The second was ‘Generativity’; a measure taken in the final testing session, when the children were asked ways of sorting a novel set of cards, to suggest ways that a boy could get to his house, and different options that they themselves had for activities after school. These were all types of activities that had been practiced during the training. The number of responses to these three questions was summed, giving a score with a mean of 8.59 (s.d. = 4.02), and a range of 1 – 16. Therefore, children who got a low ‘attempts’ score could be said to have grasped the training quickly, and children who got a high ‘generativity’ score could be said to be demonstrating flexibility across a range of situations. The MLD and ASD groups did not differ on either of these scores (Attempts ASD mean = 6.50 (2.12), MLD mean = 7.43 (4.09),  $z = -.41$ ,  $p = .68$ ; Generativity ASD mean = 7.40 (3.05), MLD mean = 10.29 (3.50),  $z = -1.57$ ,  $p = .12$ ).

The relationship between these variables and ability was calculated using Spearman’s correlations (due to the small sample sizes); table 7.11 gives the r-values. It is perhaps unsurprising, given the small size of the groups, that none of these correlations were significant; however, several of the r-values were moderate to large effect sizes. The relationship between Generativity and non-verbal ability is particularly notable for its difference in the groups – in the ASD group it is not at all related to non-verbal ability, whilst in the MLD group there is a trend towards significance ( $p=.09$ ) and a large effect size.

**Table 7.11 Spearman’s correlations between the age and ability measures and the measures of success during the EF training. ASD group in bold.**

		CA	BPVS VMA	TROG VMA	Raven’s Matrices (raw score)
Attempts	ASD (N=10)	<b>.24</b>	<b>.17</b>	<b>-.42</b>	<b>-.37</b>
(low = good)	MLD (N = 7)	-.06	-.37	-.23	-.67
Generativity	ASD (N=10)	<b>-.29</b>	<b>.29</b>	<b>.20</b>	<b>.01</b>
(high = good)	MLD (N=7)	.31	.47	.40	.68

In order to look at the relationship between success on the training and improvement on the ToM tasks, the group was split into those who improved on the proportion of ToM tasks passed, and those who did not, and the two groups were compared on the EF training variables. With the MLD and ASD groups combined,



those who improved on ToM between the pre-test and the post-test and those who did not were significantly different on both variables (Attempts,  $z = -2.46$ ,  $p < .05$ ; Generativity,  $z = -2.32$ ,  $p < .05$ ). With the groups separated the sample sizes become very small, but there is still a significant difference between the ASD groups on the Attempts measure ( $z = -1.97$ ,  $p < .05$ ) and a trend on the Generativity measure ( $z = -1.78$ ,  $p = .08$ ). In the MLD group there is a significant difference on the Generativity measures

( $z = -1.98$ ,  $p < .05$ ) and a marginal trend on the Attempts measure ( $z = -1.67$ ,  $p = .10$ ). Hence, improvements in ToM appear to be related to degree of success during the EF training. In the MLD group, those who improved ToM did not differ from those who didn't improve on age or any of the ability measures (all  $p$ -values  $> .29$ ). In the ASD group there was a trend towards a difference in the TROG ( $z = -1.74$ ,  $p = .08$ ) with those who improved on ToM having better grammar than those who did not. There were no other differences (all  $p$ -values  $> .34$ ).

To summarise, degree of success (speed of reaching criteria and a flexibility measure) during the EF training was significantly related to improvement in ToM in the EF trained groups.

### **Who improves in ToM?**

Some children obviously benefited from the training more than others, both in the ToM and EF groups, whilst some of the controls improved without any training. In order to identify factors that might have influenced improvement, the participants were divided into those who improved in ToM between the pre-test and post-test (16 ASD, 17 MLD) and those who did not (11 ASD, 4 MLD), regardless of training groups. These groups were compared on the ability scores, using non-parametric tests due to the small sample sizes. The only significant difference was in the ASD group on the TROG ( $z = -2.60$ ,  $p < .01$ ), with those who did not improve on ToM having significantly lower TROG scores than those who did improve. All other  $p$ -values were greater than .24. Improving and not improving groups did not differ on Card Sort, Trails or Go-NoGo performance at pre-test.

### **Questionnaires**

Teachers were asked to complete questionnaires about the children's behaviour before and after training (at the time of the follow-up). The results of the

‘before’ questionnaires will be discussed in detail in chapter 9. Due to a fall off in teachers returning questionnaires, only a reduced number of the children had questionnaires for both before and after (22 of the ASD group, 8 in the ToM group, 9 in the EF group and 5 controls; 11 of the MLD group, 3 in the ToM group, 3 in the EF group and 5 controls).

Table 7.12 gives the mean scores on the 3 questionnaire subscales for each group, before training and at follow-up. Two-tailed Wilcoxon’s Signed ranks test (non-parametric because of the small sample size) found a trend towards a difference between the pre-test and post-test on the ToM subscale in the ASD-Control group only ( $z = -.183$ ,  $p=.07$ ). This was towards them getting worse between pre-test and post-test. All the other differences were non-significant (all  $p$ -values  $>.18$ ). In order to look at the numbers who improved in each group, a difference score was calculated by subtracting the score at pre-test on each questionnaire from the score at follow-up. The participants were then divided into groups according to who improved between the two questionnaire time points and those who did not. The percentage and number of participants who improved in each scale in each group are given in table 7.13

**Table 7.12. Mean (s.d) scores on the 3 questionnaire subscales for each experimental group**

	ASD- ToM (N= 8)	ASD- EF (N=9)	ASD – Control (N=5)	MLD – ToM (N=3)	MLD – EF (N=3)	MLD – Control (N=3)
ToM scale (pre-test)	13.75 (4.43)	12.44 (5.27)	12.60 (5.73)	17.00 (6.25)	17.67 (2.31)	12.80 (6.02)
ToM scale (follow-up)	12.88 (2.90)	13.00 (5.39)	9.80 (3.70)	17.67 (7.02)	18.33 (4.04)	15.00 (1.58)
EF scale (pre-test)	13.50 (4.72)	11.67 (4.42)	9.60 (2.70)	14.33 (2.08)	16.00 (5.00)	11.60 (3.36)
EF scale (follow-up)	11.50 (4.99)	9.89 (4.14)	10.40 (2.97)	13.00 (1.00)	15.67 (4.62)	10.80 (4.66)
G-S scale (pre-test)	6.38 (0.92)	7.22 (1.09)	6.20 (1.30)	6.00 (1.00)	7.67 (1.53)	5.00 (0.71)
G-S scale (follow-up)	6.25 (1.04)	7.11 (1.83)	6.20 (1.92)	7.00 (1.00)	7.67 (1.53)	5.20 (0.84)

Numbers in the MLD group are too small to make any real conclusions, but there does not appear to be a relationship between being in a trained group and improving on any of the questionnaire scales. However, it is interesting that a

majority of the MLD-Control group, (who improved so dramatically between pre-test and post-test on the experimental ToM tasks), also improved on the teacher ratings of every day ToM.

**Table 7.13** *Percentage (number) improving on each scale in the experimental groups.*

	ASD- ToM (N= 8)	ASD- EF (N=9)	ASD- Control (N=5)	MLD- ToM (N=3)	MLD- EF (N=3)	MLD- Control (N=3)
Improved on ToM scale	50% (4)	56% (5)	0% (0)	67% (2)	33% (1)	80% (4)
Improved on EF scale	50% (4)	33% (3)	40% (2)	0% (0)	33% (1)	20% (1)
Improved on G-S scale	25% (2)	22% (2)	40% (2)	67% (2)	0% (0)	20% (1)

In the ASD group, about 50% of both the ToM and EF trained groups improved between pre-test and follow-up on the ToM scale, whilst none of the controls did. Fisher's exact test was done to test for a significant difference between the groups, there was a marginal trend between the ASD-ToM and the ASD-Controls ( $p = .10$ ) and a trend towards a difference between the ASD-EF and the ASD-Controls ( $p=.06$ ).

## **Discussion**

I will start by discussing the direct effects of each training programme and their implications, and will then move on to discuss the indirect effects of training and the implications of these.

### **Theory of Mind training**

In line with my predictions and with a growing number of studies (Hadwin et al., 1996; McGregor et al., 1998a; Swettenham, 1996; Swettenham et al., 1996; Wellman, Baron-Cohen et al., 2001), this study demonstrated that children with ASD (and those with non-autistic learning difficulties who fail FB) can be taught to pass standard ToM tasks. This improvement was still evident 6-12 weeks later, also in line with other studies that have assessed the maintenance of such learning over time (Hadwin et al., 1996; Swettenham, 1996). All training scenarios were supported by the 'photo in the head' strategy, and therefore the post-test and follow-up tasks differed from the training tasks in this important particular. It is therefore more likely

that children in this study were genuinely learning to use a strategy for themselves, rather than learning a fixed response from feedback, as may have been the case in some earlier studies. Generalisation to other ToM tasks was more evident in the MLD group than in the ASD group, also in line with predictions. This study differs from earlier work in a number of particulars. A wide range of ToM tasks was included in the study both at pre-test and after training, and some generalisation of learning to other tasks in the ASD and MLD groups was evident, particularly on the Penny Hiding deception task. This is the first time, to my knowledge, that this task has been used as a transfer task in training studies. However, it is not hard to see how the 'thought-picture' strategy might help to improve performance on this task, since the training focuses on how seeing something leads to knowing about it, and this is a crucial element in passing the penny hiding task, where the child must hide the penny without the experimenter seeing. The very different context of the penny hiding task (as compared to the trained unexpected transfer task) may indicate that the children are genuinely internalising the taught strategy to some extent. The ASD-ToM group also showed some improvement on the Eyes task at follow-up (70% of the group improved on this task as compared to only 14% of controls), in which it is harder to see how a 'photo in the head' strategy might help performance. It is possible that drawing the children's attention to the existence of mental states results in an increased interest in facial expressions and greater efforts to interpret the mental states behind them, however, this is entirely speculative. Other, more frequently used, transfer tasks such as the Deceptive Box (self and other) were also included, but there was no evidence of transfer of learning to these tasks in the ASD group (although there was in the MLD group). Strenuous efforts were made to minimise children's repeated exposure to FB tasks, and there were no occasions during the training in which children were given feedback on a standard unexpected transfer FB task. This was to minimise the chances of the children simply learning a strategy by rote, which previous studies have found to be effective in improving performance on the trained task, but which have not shown any generalisation in children with ASD (Swettenham, 1996).

The MLD-control group in this study proved something of a puzzle. They showed a significant improvement on the standard FB task both at post-test and follow-up, despite no access to training. This group had the highest mean pre-test ToM score (4/7), and it is possible that for some reason the unexpected transfer FB

task used at pre-test was particularly difficult for this group (it was illustrated as opposed to acted out with dolls, as were both the tasks used after training). This does highlight the need for control groups in this type of study, as it appears that, in children with MLD particularly, performance may not be at all consistent across tasks or across time. This finding also makes interpretation of the results in the MLD group more problematic, since the MLD control group improved as much on the unexpected transfer task as did the group who had been trained on that task. However, there was more evidence of generalisation in the trained MLD groups, both of which also started off at a lower baseline, which indicates that the training may have had some effect.

As in Study 2, the False Photograph test was relatively easy for all participants, and there were no significant differences between the ASD and MLD group on this task. There was no significant improvement in any of the groups on this task, but this may have been due to ceiling effects. This lack of an association between the False photo and false belief tasks may have interesting implications. If this task is properly matched to the FB task in extraneous task factors such as language and EF as has been argued (Zaitchik, 1990), then these results indicate that the MLD group did have specific difficulties with the 'belief' element of FB tasks rather than the general processing and executive demands. However, the false photo task may well be easier than the FB task, due both to the repeated exposure that children have to photographs and videos nowadays, and to the temporal grounding provided by actually taking the picture: Mitchell and Lacohée (1991) and Charman and Lynggard (1998) found that providing a photo representing a mental state for children to post during a FB task facilitated performance, and seeing the experimenter take the photo and place it face down to develop may be similar in its facilitatory effects on the task. Slaughter (1998) found that false photo tasks were significantly easier than FB tasks in preschoolers, and that performance on the tasks was not correlated.

### **Executive Function Training**

Contrary to my predictions, there was no direct effect of the EF training on the EF tasks in either the ASD or the MLD groups. This was not due to a failure to learn the strategy in the training programme, as all children passed all the stages of the training. It may be due to an inability to use this strategy in a test situation without

support, an inability to see how the strategy might be useful in the test situation, or it is possible that the training programme simply trained the wrong strategy, and did not target the specific aspects of the trained tasks with which the children had difficulties. This training programme was deliberately strategy based (rather than simply repetitively training children on the tasks, which has been successfully used in some other studies with children with ADHD e.g. Semrun-Clikemean, Nielsen et al. 1999) because it was felt that this was both more comparable to the ToM training programme and more likely to lead to genuine improvement in EF rather than to an improvement in experimental tasks only. In retrospect, the training programme might have been more effective in improving EF performance if it *had* focused more explicitly on the trained tasks, providing opportunities to practice this task without the support provided by the experimenter during the training. However, it was felt that children with ASD would be unlikely to extract their own strategy from a training programme in which this was not made explicit. Another factor which may have led to the lack of significant improvement in the EF trained groups was that the ASD-EF groups started off with slightly (if non-significantly) better EF than the controls or ToM group – 3 of the ASD-EF group got 9 or 10 on the Card Sort aggregate score at pre-test, whilst no one from either of the other ASD groups scored more than 8. These children therefore had very little scope for improvement on this task. The participants were assigned randomly to the groups, without reference to their Card Sort performance, so this possible ceiling effect was unavoidable, if regrettable.

It is also possible that the EF tests used were not simple enough to show an improvement – including more basic tasks such as the Day-Night Stroop task, for example, might have shown an improvement in the trained groups. The Trails task was very hard for a significant subset of the group, who could not complete Trails B at all (in which it is necessary to switch between numbers and the alphabet). However, it would be premature to cast this training programme as a failure, since the children did improve on tasks during training and progressed through the stages of the programme, and it appeared to have had a significant effect on ToM performance, which will be discussed in the next section.

### **Indirect effects of training**

None of the four projected scenarios described in the introduction occurred, because the EF training did not improve performance on the EF tasks. However, there

was a significant improvement on the ToM tasks in the EF trained groups, both those with ASD and those with MLD. This improvement was qualitatively different to that seen in the ToM trained group – there was no improvement on the unexpected transfer task in the EF trained groups, instead improvement at post-test was on the Penny-Hiding task in the ASD group only, and at follow-up on the Deceptive box-self in both ASD and MLD groups, and on the Deceptive box-other, and Seeing Leads to Knowing in the ASD group only. The ASD-EF group also got significantly worse on one task – the Know/guess-self task. The Knowing/guessing tasks are not conventional ToM tasks – the child must distinguish between whether someone (either themselves or the experimenter) knows what is in a box or must guess. The question is forced choice between knowing and guessing. Participants must understand the term ‘guess’ as well as ‘know’ (which the training did not involve in any form) in order to pass the task, and there is a 50% chance of answering each question correctly without any genuine understanding of mind. Performance on these tasks is therefore more likely to reflect a substantial element of chance, as compared to the standard FB tasks or the penny-hiding task, where lacking a ToM is likely to lead to consistent task failure. No other groups show a significant improvement or got worse on this task.

There was a much greater generalisation at follow-up in the ASD-EF group than that seen in the ASD-ToM group, who in the final test stage improved on the trained task only. In addition to the improvement in the EF trained groups being on different tasks to those seen in the ToM trained groups, the trajectory of improvement over time was different. At immediate post-test there was no significant improvement in the overall proportion of ToM tasks passed (although there was improvement on an individual task in the ASD-EF group, see above), however, by follow-up there had been a further improvement which was both a significant improvement as compared to the pre-test, and when compared to the post-test. In the ToM group, there was a significant improvement at post-test that was then attenuated at follow-up. This different trajectory of improvement may reflect the differing impact of the training programmes. The ToM training programme had an immediate impact on ToM performance, which is unsurprising given that it targeted those tasks specifically. The attenuation at follow-up suggests that the children may be using a strategy which some of them then forget over the next couple of months. In contrast, the EF training has only a small impact at immediate post-test, but significant differences are evident

at follow-up. It may be that because the effects of the EF training on ToM performance are more indirect, a ‘trickle-down’ effect is occurring, where the children are gradually processing the information learnt in the EF training programme, and learning to apply it in ways which benefit their ToM performance. It is possible, for example, that the set shifting which the children were trained on in EF training programme enabled them to begin to make use of their experiences of seeing different perspectives, and that this in turn lead to improvements in their ToM. It is also possible that this could be linguistically mediated – improved set shifting may enable children to begin to use their language more flexibly and to therefore use their grammatical knowledge to construct some understanding of mind. It is striking that, in the EF trained group, the only difference between those who improved in ToM and those who didn’t was a trend towards those with ASD who improved having better grammar. This finding is echoed in the whole ASD group, where those who improved in ToM were significantly better at grammar than those who failed to improve, regardless of training group. Once again, it seems that grammar is crucial for children with ASD in developing an understanding of mind, and that it may be a mediating factor when training ToM. In the MLD group this was not the case, and the EF training was in fact less effective in this group – although they improved at follow-up the improvement was in fewer tasks than in the ASD group. This is in contrast to the ToM training, which generalised more in the MLD than the ASD group.

The improvement in ToM in the EF trained groups was unexpected because the EF trained groups did not improve on the EF trained tasks, and therefore it might have been assumed that the training had failed to work. There are a number of possibilities as to why the EF trained groups may have improved on ToM tasks. Perhaps the most obvious explanation is non-specific training effects on ToM. It is possible that one-to-one sessions with the experimenter over a series of days is what improved performance on ToM tasks in both the ToM trained and the EF trained groups, and that it did not matter very much what was actually taught to the children in that time. Against this is that other non-ToM training programmes in the past have failed to improve performance on FB tasks. For example, Hadwin et al. (1996) trained children with ASD on Emotion and Pretend Play, neither of which improved performance on belief tasks. Similarly, Steerneman et al. (1996) found no improvement on ToM tasks in a socially anxious group trained with a cognitive-behavioural social skills programme. In addition, there seems no reason why non-



specific training effects should particularly affect ToM tasks whilst not affecting EF tasks. In addition the profile of improvement in the EF group was quite different to that seen in the ToM group, indicating that different factors may be at work. Finally, and perhaps most convincingly, those who improved on ToM differed significantly from those who didn't on measures of success during the EF training. It seems likely, therefore, that something specific about the training programme is affecting the children's ToM performance.

If this is the case, what might it be about the EF training programme that improved ToM without improving performance on EF tasks? There was no reference in the EF training to mistaken belief, or even to true belief, to knowledge or to any kind of mental state. The training focused on the concrete metaphor 'brain tools' which the child had to physically change in order to shift between sets, and around a 'Stop-Change-Go' traffic light sequence to aid inhibitory control and remind the children stop using one sorting strategy and shift to another. The child's attention was therefore drawn to the existence of the brain, but only in terms of its flexibility and usefulness as a tool. This training programme was focused on the child's *own* brain, in contrast to the ToM training programme which focused almost exclusively on the content of other's minds. It may be significant, therefore, that the Deceptive box-*self* task is the only one on which both the MLD and ASD EF trained groups improved at follow-up.

Exactly how the EF training affected performance on ToM without significantly affecting performance on the EF tasks is intriguing. There are three basic possibilities: (i) The EF training may be inadvertently teaching ToM rather than EF; (ii) the EF training might be boosting the children's EF skills sufficiently to overcome executive difficulties in ToM tasks (which may have been impairing their performance), but not sufficiently to improve EF skills to the extent that we would see an improvement in EF tasks; (iii) at a more developmental level, it is possible that an improvement in EF is necessary for ToM development, and the strategies which the children learnt during the training were sufficient to enable them to make leaps in ToM understanding (although it seems unlikely that such short term training as this would have a developmental impact on understanding), or, a related suggestion, that both ToM and EF may be sharing common cognitive structures at this point in development which the EF training teaches as effectively as the ToM training.

However, if both shared common structures, we would surely expect to see an improvement on EF tasks as well as on the ToM measures.

It is possible that the EF training programme was inadvertently training ToM. In particular, the ‘real life’ examples of flexibility (which included scenarios such as a child getting their brain tool ‘stuck’ and therefore going to the wrong place out of habit) had elements of FB – the brain tool was stuck because the child had failed to acquire a new belief in place of the old one. The difference was that brain tools were couched in terms of actions rather than mental states. It is possible that this might be a particularly effective way to help children with ASD think about mental states, since actions are something that they have direct experience of and can easily understand – Swettenham et al (1996) found that whilst children with ASD could use the photo analogy to predict behaviour, they could not use it to answer questions about beliefs. If we consider how the children might be using the ‘brain tool’ strategy to pass FB, it may be that they were able to pass tests such as the Deceptive box-self by considering their previous state of mind as a previous brain tool – which they have now changed, but which they can still reflect upon due to the concrete nature of the brain tool strategy. However, if they are genuinely doing this it is a huge leap from the card sorting tasks used in the training programme, and it seems strange that a training programme which is so far removed from FB tasks could effect an improvement across a range of tasks whilst the ToM programme, which was so much more closely related to mental states, did not improve performance on transfer tasks such as the Deceptive Box in the ASD group.

Alternatively, it is possible that a slight improvement on EF enabled the children to overcome executive problems on the tasks that therefore enabled them to pass ToM tasks. This would imply that EF problems had previously been masking a competence in ToM in both groups, but particularly in the ASD group who improved across a wider range of tasks than did the MLD group. From my earlier literature review on the relationship between EF and ToM (chapter 4), it does not appear likely that this is the case in children with ASD, and there is a much less clear case for a relationship between ToM and EF in children with MLD. It is also the case that there was no correlation between ToM and EF in this sample before they embarked on the training programmes (Study 2; Chapter 5). The specific tasks on which the children improved may be informative here, as ToM tasks differ in their executive demands. Both EF trained groups improved at follow-up on the Deceptive Box-self task, and the

ASD group also improved on the Deceptive Box-other and the Seeing Leads to Knowing. Since the task demands for the 'self' and 'other' sections of these tasks are virtually identical, it seems that if the improvement in task performance was due to overcoming executive problems on the task, then there should be a similar magnitude of improvement on both halves of the task – but in the MLD group, there is an improvement on the Deceptive Box-self but not Deceptive box-other. In addition, the 'Seeing Leads to Knowing' task on which the ASD group improved does not appear to have a significant executive component. There is no obvious response that is more salient than the other, as the child is being asked to choose between two present toys, rather than to repress their knowledge of reality in favour of a mental state.

The third possibility is that EF and ToM are inter-related to the extent that training children in EF will necessarily improve their ToM. For this we have to assume that the training is in some way improving EF, even though this was not evident in the EF post-tests. How this might be working is not clear. Russell (1996) suggests that self action monitoring (and therefore very early EF) is necessary for the development of a ToM, and it is possible that these children have difficulties in action monitoring which have been helped by the training programme. This fits with the finding that it is the 'self' tasks that appeared to be more susceptible to improvement than the 'other' tasks, as action monitoring is presumably initially a matter of self-awareness. However, more recent work by Russell and colleagues indicates that self monitoring is in fact unimpaired in ASD, and thus it seems unlikely that it is the cause of any ToM deficits (Russell, 2002). Russell now suggests that executive problems in ASD are a result of a failure to follow and shift between arbitrary rules or cognitive frameworks, such as sorting cards by colour rather than shape in the WCST. He suggests that this might be linked to mentalising because every day thinking involves moving between cognitive frames, and an inability to do so would constrain thinking and thus make the first person experience of thinking highly abnormal in ASD. The EF training in this task did train children to shift between arbitrary rules, and it is theoretically possible that it might be helping them to learn to shift between cognitive frameworks, and therefore improve their mentalising abilities. However, it is equally possible, as Perner (1998) argues, that ToM is an integral part of EF, and that training on EF helped ToM because in fact the elements of EF which the training programme targeted were so close to ToM. Perner argues that inhibiting undesirable action

schemas requires the representation of these schemas as representations with causal efficacy, and that this representation is in fact meta-representational and therefore ToM. The EF training programme certainly encouraged the representation of schemas – making them into concrete ‘brain tools’ that influenced action and therefore had causal efficacy. It is possible that learning about this form of representation enabled the children to begin to meta-represent. It is therefore not possible from this study to distinguish between Russell’s (e.g. Russell, 2002) and Perner’s (e.g. Perner & Lang, 2000) theories.

In addition, both of these latter options leave us with the open question of why the children did not improve on the EF tasks if they were learning to represent causal schemas as representations, or shift between cognitive frameworks, which is presumably what is needed for tasks such as the Card Sort. It may be that the tasks are of differing levels of difficulty – certainly the EF tasks used would be hard for TD 4-year-olds, whilst the ToM tasks are developmentally appropriate for that age group. In addition, the EF tasks require the child to demonstrate their abilities repeatedly (e.g. the multiple trials of the Card Sort), or to maintain them over a period of time (e.g. sustaining inhibition on Go-NoGo), whilst the ToM tasks only require a single trial. As others have argued (e.g. Leslie & Polizzi, 1998) a true control for the executive demands of a ToM task would be a single trial of an EF task, not repeated trials. It may be that limitations of the tasks also account for the lack of a direct effect of training – variation in performance between testing sessions was wide in all of the experimental groups, and it is possible that these EF tasks were simply a less stable measure than the ToM tasks in these groups, and that therefore it was much harder to discern any improvement.

Indirect effects in the opposite direction were far less marked. There was a slight tendency on the part of the ASD-ToM group towards improvement on the EF tasks, however, this only reached significance in the Card Sort, where the ASD-Controls group also improved. It therefore seems unlikely that this was directly to do with the ToM training – unlike the MLD-Controls, there is no reason to think that the ASD-Controls were under performing on the pre-test. In the MLD group there were no effects at all of ToM training on EF, and the MLD-ToM group were the most responsive to the ToM training, demonstrating generalisation to untrained tasks both at post-test and at follow-up. It therefore seems that this type of ToM training did not improve performance on these EF tasks in children with MLD or with ASD.

### **Responses in the ASD and MLD groups**

Part of the reason for including both ASD and MLD participants in this study was to compare the effects of training in the two groups, with the aim of investigating possible differences in difficulties underlying FB failure in these groups. The hypothesis was that the MLD group's difficulties would be due to factors other than genuine problems in ToM, and that therefore training them might be particularly effective, and also that EF training might selectively improve their ToM performance (as opposed to the ToM performance of the ASD group). These predictions were partially supported. The ToM training was particularly effective in the MLD group, not only did the children learn to pass the trained unexpected transfer task but they also showed generalisation to both the Penny Hiding deception task and the Deceptive box-other. This was wider than the generalisation seen in the ASD group, and was maintained at follow-up. This is in line with other studies, which have found that young TD children and those with non-autistic learning difficulties showed a greater degree of transfer to other ToM tasks than did children with ASD (McGregor et al., 1998a; Swettenham, 1996). However, the EF training improved ToM performance in both the ASD and MLD groups, and in fact appeared to be particularly effective in the ASD group, which goes against my prediction. As I have discussed above, there are several possibilities as to why the EF training might have had this effect on ToM performance, but there is no evidence from this study to suggest that it was more effective for participants with MLD.

The group that really stands out as being quite different to other participants in this study was the MLD-control group. They were the highest achieving group on ToM tasks at the pre-test and showed a significant improvement on unexpected transfer tasks both at post-test and at follow-up. Whilst it seems likely that the training effects seen in the other MLD groups are genuine, since they mirror those seen in the ASD group and also show a degree of generalisation not seen in the controls, the performance of the MLD controls serves as a warning to studies that do not include control samples, or who assume that performance on a single FB task indicates a stable cognitive ability. Evidence from my previous study of ToM and language (Study 1: Chapter 3) as well as from this study indicates that performance on FB tasks in children with MLD may be particularly vulnerable to task factors. However, the fact that the changes in MLD trained groups mirrored the pattern of improvements seen in the ASD group suggests that it would be premature to conclude

that children with MLD's difficulties with ToM tasks are entirely different in nature to those of children with ASD. Taken with the results of Study 2 (Chapter 5), which found a particularly high alpha value between ToM tasks in children with MLD, it may be that some children with MLD do have difficulties in representing other minds, and that these difficulties are not due simply to difficulties in reflecting on representations in general (as they had no problems on the false photograph task), nor are they due to executive problems (as executive tasks did not relate to performance on FB tasks in Study 2). However, these difficulties are clearly quite different to those seen in people with ASD, both in the proportion of individuals affected (see Study 1: Chapter 3) and in the severity of these problems – children with MLD who failed FB tasks performed much better on a range of other ToM tasks than children with ASD (Study 2: Chapter 5).

### **Real life measurements**

The 'before' and 'after' questionnaires for teachers (who were blind to group assignment) were completed over a relatively short time scale (in most cases 2 months), and it seemed unlikely that any changes in real life behaviour would be seen over this time, particularly since the same teachers were completing the questionnaires and so would probably have quite a strong idea of how they thought a particular child behaved. However, there was some tentative indication in the ASD group only that those who were trained were more likely to improve on the ToM scale, with over half of the trained children (50% of the ToM trained groups, 56% of the EF trained group) showing some improvement over time, as compared to none of the control group. However, a much larger sample size is needed over a longer period of time in order to really investigate this issue. Study 5 (Chapter 9) will examine the relationship between performance on ToM tasks and real life behaviour in more depth.

Numbers with questionnaire scores both before and after were so small in the MLD group that it is not possible to make any conclusions, but there was no evidence for a relationship between improving on experimental tasks and improving on the questionnaire scales. However, a high proportion of the controls (who improved on experimental tasks), also improved on the real life ratings.

## **Conclusion**

In conclusion, this study shows that after a relatively brief intervention programme, children with ASD and MLD are able to learn to pass FB tasks and to show some degree of generalisation to other, non-trained tasks. This learning was sustained at follow-up. There was tentative evidence from this study that in those with ASD some cases both the EF and ToM training programmes were associated with a real life improvement on behaviours thought to require a ToM. The success of this intervention in stimulating some generalisation of learning relative to other studies may be due to the use of an alternative strategy for mentalising (as in Swettenham et al. 1996), as opposed to using straightforward feedback on FB tasks (Swettenham, 1996) or a rule based learning procedure (Hadwin et al, 1996). Children with MLD showed more generalisation to other ToM tasks than children with ASD. This may indicate that the ToM difficulties demonstrated by the children with MLD were less profound than those seen in the ASD group, and were thus more easily overcome, or it may indicate that the MLD group were simply better at generalising. Given that children with ASD are known to have difficulties in generalisation (e.g. Plaisted, 2001), but also that studies 2 and 5 of this thesis indicate that children with MLD who fail FB tasks do have less profound problems in social interaction than do children with ASD, some combination of the two may account for the superior improvement of the MLD group. This superior generalisation was also in line with previous studies, which have found some transfer of ToM learning with children with DS (Swettenham, 1996), and with social anxiety and MLD (Steerneman et al., 1996).

The trajectory of improvement in the two diagnostic groups was very similar, somewhat unexpectedly. This may indicate that there are some similarities between children with ASD and those with MLD who fail FB tasks, and that the ToM difficulties of those with MLD should not be dismissed as artefactual or irrelevant, although these difficulties do appear to be both less pervasive and less severe than those seen in children with ASD (Studies 1 and 2, chapters 3 and 5).

Training in EF improved children's performance on a range of ToM tasks, despite not improving their performance on EF tasks. This may be because the training programme itself inadvertently trained some aspects of ToM, or because EF is causally related to ToM, either as a precursor or because ToM is an integral part of EF. This indirect transfer of learning from the EF training to ToM was evident in both the ASD and MLD groups, with the ASD group showing a wider degree of

generalisation than did the MLD group. Group differences, therefore, showed the reverse pattern to that seen in the ToM trained groups, and quite why that might be is something of a puzzle. It may be that the EF training programme inadvertently targeted some aspect of EF that children with ASD find hard about ToM tasks, and harder than do children with MLD. Possibilities could be flexibility or inhibitory control (although again we have the problem that there was no improvement on tasks measuring these abilities). These suggestions are purely speculative at this stage.

There are a number of unanswered questions from this study. However, it does appear to indicate that children with both ASD and MLD can be taught over a short-term intervention programme to pass FB tasks, and that this can generalise to untrained tasks in at least some children. This leads us to ask what exactly it means for a child with ASD or MLD to pass, rather than fail, a FB task, and whether it has any implications for their social interactions in everyday life. Up to this point this thesis has focused on failure on FB tasks, and what that might mean in children with ASD or MLD. The focus will now shift to look at those who pass FB tasks – a minority of children with ASD, but a majority of those with MLD. This group of children will be investigated in greater depth in chapter 8, and the everyday social behaviour of both FB passers and failers will be examined in chapter 9.



## **CHAPTER 8**

### **STUDY 4: WHAT DOES FB TASK SUCCESS MEAN IN INDIVIDUALS WITH ASD?**

#### **Introduction**

As discussed in chapters 1 and 3, a certain proportion of individuals with ASD pass first order FB tasks (e.g. Baron-Cohen et al., 1985; Happé, 1995), and some of those also pass second order FB tasks (e.g. Bowler, 1992). Despite this, there has so far been relatively little investigation of what it actually means for people with ASD to demonstrate an understanding of FB, and what it might be that enables them to do so. It is possible that passing FB tasks in ASD demonstrates little more than problem solving ability, ‘hacking’ out a solution with no genuine understanding of mind (Frith et al., 1991). Alternatively, this group may genuinely have better socio-cognitive skills than the majority of children with ASD, in which case we would expect this to be demonstrated in other experimental tasks and in real life. These better socio-cognitive skills may well not be achieved in the same way as TD children develop their skills, however, and the relationships between ToM tasks, verbal and non-verbal ability may inform us as to whether there are differences in the way that individuals with ASD solve ToM tasks as compared to individuals with MLD.

There are some clues as to the characteristics of this group. Happé (1995) identified high verbal ability as a shared characteristic of individuals with ASD who pass FB tasks, and Study 1 of this thesis found that receptive grammatical ability was a particularly good predictor of performance on FB tasks in children with ASD, perhaps suggesting that good language skills are necessary for the development of ToM, rather than vice versa. There is also evidence that, at least in some cases, superior real-life socio-cognitive strengths are associated with passing FB tasks. Frith et al. (1994) found superior performance on teacher ratings of social behaviour requiring some consideration of mental states in individuals with ASD who passed FB tasks as compared to those who failed, indicating that passing a FB task does indicate some degree of competency in social behaviour in everyday life. The present chapter focuses on the performance of children with ASD who pass FB tasks on a range of advanced tests of ToM, with the aim of identifying strategies they may be using, and of identifying features which make a ToM task easier for individuals with ASD. The

specific issue of the relationship of everyday life behaviour to performance on ToM tasks will be addressed in study 5 (chapter 9).

Experimentally, differences between those who pass FB tasks and those who fail have been found on experimental tasks such as the Strange Stories, in which participants are asked questions requiring the attribution of a mental state to the protagonist in a short story (Happé, 1994). Happé (1994) compared three groups of individuals with ASD, those who failed FB, those who passed 1<sup>st</sup> order FB only, and those who passed 2<sup>nd</sup> order FB as well as 1<sup>st</sup> order. Both groups of passers outperformed the failers, and those who passed 2<sup>nd</sup> order FB outperformed those who passed 1<sup>st</sup> order. This group who passed 2<sup>nd</sup> order FB still performed worse on the Strange Stories task than normal adult controls, however. They performed indistinguishably from young TD controls and controls with MLD, despite having a large verbal IQ advantage. It is also true that very able individuals with ASD (who in most cases will pass at least 1<sup>st</sup> order FB tasks, although this is not assessed in every study) still have difficulties in other advanced ToM tasks such as recognising social faux pas (Baron-Cohen, O'Riordan, Stone, Jones, & Plaisted, 1999), recognising mental states from the eye regions of the face (Baron-Cohen, Wheelwright, Stone, & Rutherford, 1999) and spontaneous social reasoning about animated shapes (Abell, Happe, & Frith, 2000; Klin, Schultz, & Cohen, 2000). All the studies cited above found deficits in individuals with ASD when compared with normal adults or children, rather than to people with learning difficulties. The problem of choosing appropriate controls for high-functioning participants with ASD will be discussed later.

From the literature to date, we are left with a picture of an able group of individuals with ASD, who pass 1<sup>st</sup> order FB and in some cases 2<sup>nd</sup> order FB tasks. Some of these individuals demonstrate superior social insight in everyday life, and on other ToM tasks, when compared to individuals with ASD who fail FB tasks, yet they still show some of the social difficulties associated with ASD, and are impaired on some experimental ToM tasks when compared to normal controls. This chapter investigates the performance of this group further, in particular assessing whether there are certain experimental manipulations that make the tasks easier or harder for participants with ASD that might shed light on how they may be solving these tasks. In this introduction I will first discuss what might account for the superior

performance of this group, in particular discussing the potential role of language in the understanding of mind in ASD, and will then outline the predictions for this study.

Firstly, what might be accounting for the superior performance of these high-functioning individuals? The obvious place to start is with their shared strength, their language. Several studies (reviewed in chapter 2), and Study 1 (chapter 3) of this thesis indicate that individuals with ASD who pass FB have a higher level of verbal ability (vocabulary and grammar) than those who fail, and a higher level than children with MLD or TD children who pass FB tasks. This has led to suggestions that language itself might be facilitating FB performance in people with ASD, that they might be using their superior language skills to provide the structure for a rudimentary understanding of mind (Happé, 1995; Tager-Flusberg, 2002). Exactly how a high level of verbal ability might lead to a better ToM ability has not received much experimental investigation to date.

### **How might language enable individuals with ASD to form a better understanding of mental states?**

In chapter 2, the role of language in the typical development of ToM was discussed in detail. The case of those able individuals with ASD who appear to have some understanding of mind may be quite different, to that seen in typical development, and will be discussed below. From a reading of the literature to date it seems there are four alternative proposals for the way in which language may enable an individual with ASD to pass ToM tasks, and possibly develop a genuinely better understanding of minds. These proposals will be briefly outlined in turn.

The first option is that language may be important developmentally, as a prop to enable children with ASD to develop some degree of understanding of mental states. Its importance may be due to the pathway it provides into learning about mental states, providing access to a great deal of explicit information or rules of thumb about thoughts and beliefs. Anecdotally, people with high-functioning ASD are often observed to appear to be following learnt social rules that are sometimes even explicitly verbalised, and to have difficulties when the situations change and more flexible behaviour is required. Experimentally they have difficulties in tasks where the conventional social rules do not apply, or can be misapplied (e.g. Baron-Cohen et al., 1999; Blackshaw, Kinderman, Hare, & Hatton, 2001). If individuals with ASD are using language in order to acquire verbal rules about mental states, then

tasks that allow the participants the opportunity to access schemas of previously learned information should be relatively easy, whilst tasks that require a more immediate and fluid type of mental state processing should still be difficult.

An alternative form of this developmental hypothesis argues that the structure of language itself could itself provide individuals with ASD with the context to start to distinguish between different perspectives. As I have discussed elsewhere (chapter 2), De Villiers (e.g. De Villiers, 2000) argues that the complement syntax structure inherent in mental state verbs is crucial in coming to understand the differing perspectives of other minds. If this were to be the case, then those people with ASD who have a high enough level of language should be equally unimpaired on all ToM tasks, regardless of their language demands, since in De Villiers' account language is important as a developmental stepping stone to a ToM, rather than as a prop during the online use of a ToM.

Tager-Flusberg (2001) has a different approach to ToM which speaks directly to the issue of those with ASD who pass ToM tasks. She argues that ToM can be dissociated into two components. These are a 'social-perceptive' component which refers to the online immediate judgement of a person's mental state, based on the information available in their faces, voices and body posture, and a 'social cognitive' component, which refers to the capacity to make complex cognitive inferences about mental states, requiring the integration of information across time and events. Tager-Flusberg argues that in the case of TD children, the social cognitive component builds on the social perceptive component, but that in the case of ASD the social perceptive component is impaired. In most individuals with ASD this means that the social cognitive component will also be impaired. It is possible, however, for some high-functioning people with ASD to develop some of the social cognitive component of ToM, but they have to do it via a different developmental pathway, relying exclusively on language (Tager-Flusberg, 2000, 2002) or, alternatively, on more general logical reasoning skills to 'hack out' a solution (Frith et al., 1994). Tager-Flusberg argues that communication verbs in particular may be important in enabling individuals with ASD to draw the distinction between belief and reality and to bootstrap themselves into a better understanding of mind – i.e., whilst an individual with ASD may not understand that someone has a mistaken belief, they can understand that someone has said something mistaken and therefore they can begin to get an idea of differing perspectives. Tager-Flusberg argues that the

understanding of minds that individuals with ASD gain in this way will not be as intuitive or as fast as that of TD children, but will nevertheless enable individuals with ASD to make some accurate attributions about mental states. Tager-Flusberg (2002) used the analogy of making lists of street directions in order to get from one place to another as a compensatory strategy for a lack of spatial awareness. In the case of ToM this compensatory strategy could involve constructing a verbal narrative from the situation – ‘talking it through’ in rather the same way as a TD child might find talking through a maths problem helpful. This would lead to the prediction that tasks that have a well defined verbal structure would be relatively unimpaired, as these tasks provide a substantial element of the linguistic structure already and therefore much of the work is done, in comparison to every day life where there is no narrator, and no one to point out salient features of the situation. Tager-Flusberg’s model would suggest that whilst individuals with ASD will be able to do relatively well on social cognitive tasks (such as the Strange Stories; (Happé, 1994)), they will remain impaired on social perceptual tasks such as the Eyes task (Baron-Cohen et al., 1997).

This brings us to the fourth alternative explanation for the importance of language in ToM performance. Language may be important as an online tool, enabling the individual with ASD to solve problems involving mental state attributions by verbal means. In the example above, the use of language to construct a verbal narrative would be an online strategy, even though the ability to do so may have developed due to superior language skills (and therefore language is implicated both developmentally and online). Using learnt rules to solve tasks would also be an online strategy, although acquiring the rules would be a developmental process. It is also possible that high-functioning individuals with ASD have a non-verbal online strategy – they may be able to tap into mental models, for example, constructed through learnt rules. All theories that postulate the use of online strategies would predict slower and less accurate mental state attributions as compared to controls. Bowler (1997) tested this with a small sample of adults with ASD, recording their reaction time for 2nd order FB tasks. Although the ASD group had a longer reaction time than controls on the FB questions, this was equally evident in questions that did not involve mentalising. However, it may well be that 2<sup>nd</sup> order FB tasks (which are passed by TD 6-year-olds) are not sensitive enough to pick up these differences in able adults with ASD. Any online strategy for such relatively straightforward tasks may be so overlearnt that it is no longer notably time consuming or prone to error.

Online strategies have the added feature that they may or may not be helpful in real life – Frith et al. (1994) suggest that some individuals with ASD may use ‘hacking’, non-ToM strategies which, though they may enable individuals to solve some ToM tasks, would not be very useful in every day life. If online strategies are being used, it is likely that two groups exist - individuals with ASD who use task-specific online strategies to help them pass experimental tasks, and those individuals who use online strategies that also enable them to improve their everyday functioning. It is also possible that some individuals with ASD may be attributing mental states in a post hoc fashion, when the need becomes apparent (as it often does in ToM tasks, but maybe does not so often in real life).

This study was designed to test several related hypotheses about the way in which high-functioning individuals with ASD may be overcoming some of their difficulties in ToM tasks. The first prediction was that high-functioning children with ASD would not have difficulties on ToM tasks in which they could potentially tap into internal sources of learned information and rules. These tasks included a vocabulary test of mental states and emotion terms, and a task asking the participants to report their ability not to think (Flavell et al., 1997). Knowing that it is not possible to stop thinking appears to be something that children can learn, as scientific fact. Paired tasks were included which shared certain similarities with the above tasks, but which did not appear to be so open to rule use. These included the Children’s Eyes task (Baron-Cohen et al., in press), which used the same mental state and emotion terms as the Eyes vocabulary task, but required the child to pair those terms with a photograph of a person’s facial expression, identified only by the eye region, and a task in which children were asked to report their thoughts (Flavell et al., 1997).

The second prediction was that ToM tasks with a well-defined verbal structure would be less problematic for high-functioning individuals with ASD than tasks with no verbal narrative. The verbally structured tasks were second order FB tasks, and stories requiring mental state attributions (Happé, 1994). A range of tasks which were not thought to be soluble by learnt rules, and which did not have a verbal structure, were also included, with the prediction that impairments would remain on these. A non-verbally structured task was included, on which it was predicted that the children with ASD would have difficulties. This task involved inferring mental states from cartoon pictures (Happé et al., 1999).

The third prediction was that people with ASD would calculate mental states as and when they were required rather than as an ongoing process; in a rather ‘post-hoc’ fashion. In order to test this, a variation of a task designed by Paul Harris was developed. Harris (Harris & Martin, unpublished) designed a task that used the relative reading speed for sentences containing an emotion to test whether the reader judged the appropriateness of various emotional responses from the protagonist’s standpoint or from their own standpoint. Readers will take less time to read a sentence containing an emotion which is congruent with the story they are reading (called ‘appropriate’ throughout this account) than one containing an non-congruent (and thus ‘inappropriate’) emotion, and this can be used to assess what the reader judges to be an appropriate emotion in the case of a knowledgeable or ignorant protagonist. Harris’s aim was to investigate whether mental states were attributed throughout a story, or whether they were attributed when it became necessary to do so, and he therefore compared reading times for appropriate emotions when the protagonist had a false belief with times for when the protagonist was fully knowledgeable about the situation. In normal adults, he found that mental states were attributed throughout the story, it did not take any longer to read an emotion sentence based on a false belief than one based on a true belief. A variation of this task was designed for the children in this study, which involved stories in which the emotional content of a situation changed (e.g. Jane had thought she would be going to the cinema, but actually would be going to the dentist instead). In four stories, the protagonist was informed of the situation (e.g. Jane’s dad told her they would be going to the dentist instead), whilst in the other four they were kept ignorant (e.g. Jane’s dad had not yet told her and so Jane still thought she was going to the cinema). Each story ended with a sentence containing an emotion term about the protagonist. The sentence containing the emotion term (she/he felt happy/sad) was appropriate for half the stories (e.g. Jane was sad because she was going to the dentist), and inappropriate for the other half (e.g. Jane was happy because she was going to the dentist). The reading times for the various conditions tell us which the reader judges to be appropriate or inappropriate. For example, if Jane doesn’t know she will be going to the dentist instead of the cinema she should be happy, but if the reader fails to take account of her lack of knowledge then he will assume she would be sad as this would be congruent with the actual situation as opposed to Jane’s perspective on the situation. This task thus provides an ‘implicit’ measure of perspective taking through

reading time. In addition, if the reader is not taking the mental standpoint of the protagonist throughout the story (as did the normal adults in Harris's study), but is instead waiting for the emotion sentence and then computing the mental state in a somewhat post hoc way, reading times for the appropriate emotion in the ignorant condition (i.e. when Jane doesn't know she will be going to the dentist's) will be slower than that in the knowledgeable condition (where there is no conflict between the protagonist's and the reader's knowledge state). The prediction is that this will be the case for some members of the ASD group.

In order to test whether there was any evidence for individuals with ASD using a specifically *verbal* online strategy, a task was designed which involved interfering with the participants' inner speech through articulatory suppression. Children had to repeat the word 'baa' whilst ordering picture sequences that told a story involving a FB. In the control condition they tapped on the table instead of saying baa. Mechanical stories were also included as a control for the other (non-ToM) demands of the task. If children with ASD were using a verbally mediated strategy at the time of doing the task, then the verbal interference condition should selectively impair their ability to sequence the stories involving mental states.

There were therefore 4 main questions in this study. First, are tasks that allow access to learnt rules (versus those that don't) easier for individuals with ASD? Second, are tasks with a well-defined linguistic structure easier for individuals with ASD? Third, is there evidence that individuals with ASD are using post hoc strategies in mental state attribution? Fourth, is there evidence that strategies are verbally-mediated?

A final, subsidiary, question, and one which will be discussed further in chapter 10, is the role of executive function in ToM in individuals with ASD. A range of EF measures were not included in this study, however, a working memory task was included, in order to test the prediction that an alternative strategy for ToM would require a certain degree of 'holding in mind' which would need working memory. If this is the case, then working memory should be related to ToM performance in children with ASD, and this should be less evident in children with MLD.

Since there are no standardised scores for the ToM tasks used, there is no way of comparing results between the various tasks, and it is not possible to directly assess whether the participants with ASD find one task easier than another. We therefore



used comparisons with a control group to give an indication of which tasks the groups are impaired or unimpaired on. Selecting a control group for high-functioning people with ASD has long been regarded as problematic. Many studies have used normal adults, or TD children (e.g. Baron-Cohen et al., 1999; Blackshaw et al., 2001) as controls for high-functioning individuals with ASD. This study used individuals with MLD, since a theme of this thesis has been the comparison of the ToM performance of individuals with ASD and MLD. It is also the case that a matched sample of TD children would have been practically impossible to find. The range of verbal ability (as measured by the BPVS) in the present sample of high-functioning children with ASD went from 52 to 156 (VMA ranging from 5.92 – 17.00). Had we matched on VMA alone the TD group would have had to include individuals who were both much younger than the youngest participant with ASD (who was nearly 10 years old) and individuals who were significantly older than the oldest participant (who was 15 years old). The tasks were designed to be appropriate for the range of CAs included in the ASD group. Therefore a group of children with moderate learning difficulties of similar CAs to the participants with ASD were used, although, due to the nature of their difficulties, this group had a lower mean VMA and ability level than did the ASD group. This was inevitable given that passing ToM tasks in autism is associated with a much higher verbal ability level than in those with MLD (Happé, 1995). Since this study predicts that the MLD group should outperform the ASD group on certain tasks, this difference is less problematic than it might have been. In order to examine whether any whole group differences (or lack of differences) were due to differences in developmental level or IQ, smaller, ability matched, subgroups were created and all analyses were repeated with these groups as well as with the full sample.

## **Method**

### **Participants**

22 children with MLD and 19 children with ASD took part in this study. The children with MLD had all participated in Study 1 (Chapter 3), and were chosen for this present study because they passed at least 2/3 of the basic unexpected FB tasks described in this study. 12 of the children in the ASD group were participants in Study 1, and an extra 7 participants with ASD were also recruited. These children had participated in a previous unrelated study and were known to pass standard FB

tasks of the type administered in Study 1. Ten of the ASD group had a diagnosis of autism or ASD, whilst 9 had a diagnosis of Asperger's Syndrome. Table 8.1 gives their relative ages and ability levels. In the MLD group there were 18 boys and 4 girls, and in the ASD group there were 17 boys and 2 girls.

**Table 8.1. Age and ability measures for the two groups; mean (s.d.)**

	ASD (N = 19)	MLD (N = 22)
CA (years)	12.13 (1.37)	12.79 (0.91)
BPVS VMA (years)	12.16 (2.97)	9.05 (0.96)
BPVS standardised	102.32 (22.14)	76.55 (5.35)
TROG VMA (years)	9.18 (2.17)	7.39 (1.60)
TROG age-adjusted	134.79 (24.98)	109.22 (20.13)
Raven's CPM (raw score)	29.89 (5.44)	26.95 (5.30)

Non-parametric tests were used to compare the groups, since the variances were not homogenous. The groups differed significantly on all the ability measures (all  $p$ -values  $<.05$ ) but did not differ significantly on CA ( $p=.10$ ). Miller and Chapman (2001) argue that it is not possible to control for variables (such as ability) which differ systematically between groups by covarying out these variables using ANCOVA, particularly when the groups are not randomly assigned and when the disparity is inherent in the groups, as in this case. To check the effects of ability differences, therefore, subgroups were selected that were more closely matched on ability. The aim of these groups was to maximise the numbers included whilst creating groups who were comparable in their developmental level, as assessed by the VMA scores and the raw score on Raven's CPM. This was done by excluding the most able children in the ASD group (those with a BPVS VMA of more than 13 years) and the least able in the MLD group (those with a BPVS VMA of less than 8.5 years). Again, non-parametric tests were used to compare the two groups since the variances were not homogenous. These groups did not differ significantly on BPVS VMA ( $p=.09$ ), TROG VMA ( $p=.49$ ) or Raven's CPM ( $p=.89$ ) although they differed on the BPVS standardised score ( $p<.01$ ), the TROG age-adjusted score ( $p<.05$ ) and CA ( $p<.01$ ). All analyses were repeated with these groups in addition to using the whole sample. Table 8.2 gives the ability and age of these subgroups.

**Table 8.2. Age and ability of ability matched subgroups; mean (standard deviation)**

	ASD (N = 11)	MLD (N = 17)
CA (years)	11.54 (1.31)	12.99 (0.92)
BPVS VMA (years)	10.17 (2.09)	9.38 (0.73)
BPVS standardised	90.64 (18.37)	77.59 (5.01)
TROG VMA (years)	8.41 (2.48)	7.81 (1.55)
TROG age-adjusted	132.20 (31.45)	111.69 (21.67)
Raven's CPM (raw score)	27.45 (5.84)	27.82 (5.14)

### **Procedure**

Participants were seen in a quiet room in their school or home by a single researcher. The testing session took approximately two hours. In the case of those participants seen at school, this was split into two or three shorter sessions to fit in with school timetables. In the case of those seen at home, the session was split into two with a short break in the middle. The tasks were presented in one of two fixed orders, there were no order effects.

### **Measures**

#### **Ability measures**

**British Picture Vocabulary Scale:** This task is described in Study 1 (Chapter 3).

**Test of Reception of Grammar:** This task is described in Study 1 (Chapter 3).

**Raven's Coloured Progressive Matrices:** This task is described in Study 2 (Chapter 5).

**Working memory: Digit Span:** The forwards and backwards digit span tasks from the WISC (Wechsler, 1974) were administered.

The forwards digit span measures the number of digits the child can repeat back in order, and taps into the phonological loop of the working memory system.

The backwards digit span requires the participant to repeat back digits in order, and so taps into the working memory element of the central executive.

### **Theory of Mind measures**

Table 8.3 gives a summary of the tasks used in this study, with the area that they were designed to test. The tasks are described in detail underneath.

***Table 8.3 A summary of the tasks used in Study 4, and the reason for their inclusion***

Eyes vocabulary test (vs. Eyes task) Not-think task (vs. Think task)	Tasks which could be solved by learnt information (paired task which could not be solved with rules)
2 <sup>nd</sup> order FB Strange Stories (vs. ToM cartoons)	Tasks with a well-defined linguistic structure (paired task without a linguistic structure)
Unexpected Stories	Tasks designed to assess post-hoc mental states attributions
Picture sequences	Tasks designed to interfere with verbal online strategies

Children's Version of the 'Reading the Mind in the Eyes' task (Baron-Cohen et al., in press): This task was included as a task on which the participants could not use learnt rules. In addition, it is a task without a well defined linguistic structure. It is described in Study 2. All 28 items were administered to the participants in this study. Scores could therefore range from 0-28.

Eyes vocabulary test: This task was designed to check that the children understood all the words used as correct answers in the Eyes task, and also to give a contrast task for the Eyes tasks that could be solved using learnt rules or information (and was therefore more of a social cognitive than a social perceptual task). Therefore a new vocabulary test was devised, using the correct answers from the Reading the Mind in the Eyes task. Participants were asked what a series of words meant, having first been told they could explain what the word means, give another word which means the same thing, or describe a time when they had felt like that. An example was given as follows. *'If I asked you 'What does angry mean? you could say 'It means when you are very cross because someone has done something you do not like.' And for an example, you could say, 'I got very angry when my brother kicked me, because he hurt me''.*

Participants were asked what 16 words meant. There were fewer words than items on the Eyes task because of repetitions of some words within the Eyes task. These words were: jealous, thinking, remembering, friendly, relaxed, worried, interested, sad, kind, believe, happy, serious, nervous, hoping, sure and pleased.

Their answers were coded as follows: 2-points were awarded for a correct definition, an appropriate synonym (e.g. miserable for 'sad'), or an example which made it clear that the child understood the word (e.g. we were sad when my granddad died because we missed him). 1-point was awarded for a partial definition, appropriate use in a sentence that did not completely clarify the meaning of the word (e.g. you remember to go out to play when the bell rings), or a definition that included some inaccuracy as well as some accuracy. 0-points were awarded for a wrong answer, a non-informative or simply repetitive sentence (e.g. thinking is when you think about something) or 'do not know'. These scores were summed to give a total score that could range potentially from 0-32. Twenty randomly selected scripts (49%) were double coded by two independent raters. Kendall's tau-b was .76 for the task, indicating a very good level of agreement.

Think/Not-think task: These tasks were included as quite a different measure of the understanding of mind than those usually used with participants with autism, and as measures of introspective ability. They were taken from Flavell (1992). The tasks started with a short pre-training to get the child used to the procedure. In this, the child was first asked to say the first thing that they thought in response to a word spoken by the experimenter. They were then asked to close their eyes, think of two things that a spoken word made them think of, but not to say anything until they were asked by the experimenter. They were then told that people sometimes have thoughts they do not try to have, and sometimes thoughts happen so fast that it is hard to notice or remember them. The Think task was then administered. The participant was asked to move to a chair or corner labelled 'think' and told to close their eyes and think about something they would really like to do. Following 5-10 seconds of silence they were asked to think about something they really did not like to do. Following another period of silence they were asked to return to their original chair and were asked to tell the experimenter all the thoughts they had had. Once they had done this they were asked if they had any other thoughts, and then were asked to recount their thoughts one by one (the experimenter asked 'what was your first/second/third etc thought?' until the child said they had had no further thoughts). They were finally asked if they had thought about what they had had for breakfast, as a measure of their suggestibility.

The Not-think task was then administered. Participants sat in a chair or corner labelled 'Do not think', closed their eyes and were asked to try not to think of

anything at all, to try to keep their mind completely empty of thoughts. After 20-25 seconds they were told to open their eyes and move back to their original chair. They were asked about their thoughts during their time in the 'Do not Think' chair. If they said they had had some thoughts they were asked to recount their thoughts in order (the experimenter asked 'what was your first/second/third etc thought?' until the child claimed to have had no more thoughts). They were then asked if it had been easy or hard to stop themselves having thoughts (with the order of easy/hard counter-balanced). They were then asked why and what they had done to try and stop themselves having any thoughts.

Scoring followed a five point scale for each task. In the 'Think' task, participants were given one point for recounting both a like and a dislike, one point for placing the 'like' thought first in their recollection (either spontaneously or when asked to recount their thoughts one by one), one point if they showed evidence of genuine recollection of visualisation and imagination rather than simply saying something they liked or disliked (e.g. saying 'I thought about being a professional football player and winning the cup, holding it up high and feeling great' would get a point, whereas 'Football' would not). They got another point if they used mental state terms other than repeating the phraseology used by the experimenter, and one point if they *did not* say they had thought about what they had for breakfast. These five points were then summed to give a total score ranging from 0-5.

The 'Not-think' task was scored as follows. Two points were awarded if a child said they had had thoughts and reported some content. A single point was awarded if the child either said they had thoughts but were unable to report their contents, or if they denied having thoughts but then reported some (e.g. if they said 'No I did not have any thoughts, I just thought of a black room'.). A further point was awarded if they said it had been hard to stop themselves having thoughts, another point if they reported a mental strategy for doing so, and a final point if they gave an appropriate explanation of why it was hard to stop themselves having thoughts (e.g. because your thoughts just keep going, or because you can't stop yourself thinking, your brain keeps working). These points were summed to give a total 'not think' score out of five.

Twenty (49%) of the scripts (10 from each group) were double scored by an independent rater. Kappa's for the Think task ranged from .86 –1.0 and from .76-1.0 for the Not-think task indicating an excellent level of agreement.

2<sup>nd</sup> order FB: These were included as standard tasks used to assess advanced ToM, and as an example of ToM tasks with a well-defined linguistic structure. Two 2<sup>nd</sup> order FB tasks were administered, illustrated by line drawings. These tasks were based on Baron-Cohen (1989) and Bowler (1992). Both tasks included a 1st order question, a 2nd order question and a 2nd order justification question. All participants passed the 1st order question, and so scores were calculated as follows: Participants were given a point for passing the 2<sup>nd</sup> order FB question, and an extra point if they also passed the corresponding justification question. These two scores were summed, giving a potential range of scores of 0-4 for the FB tasks. Participants were not given credit for their performance on the justification question if they failed the FB question. An example script is given in Appendix VI.

Strange Stories (Happé, 1994): The experimenter read six of the children's version of the Strange Stories out loud to the child, each of which was followed by a test question. The stories were left in front of the child so they could read them again if they wished. Three stories (the ToM stories) involved mental state attributions (persuasion, double bluff and deception) and three involved deductions about physical events (the Physical stories). Example stories (to which the example answers below refer) are given in Appendix VI. The ToM and Physical stories were matched in difficulty from the results of a sample of 52 TD eight-year-olds (Fisher et al., submitted).

For both types of story, answers were coded on a 0-2 scale. The maximum score of 2-points for a story indicated a full answer, in which the child explicitly made the correct inference about events (e.g. for a ToM story - *Because he's a liar but he told him there; he knows that he is a liar and he done it*; for a non-ToM story - *Because the furry thing was in there and he touched the beam which made the alarm go off*). 1-point was given for answers which referred to the correct facts, but were not explicit (e.g. for a ToM story *Because he lies*; for a non-ToM story *Because the animal set the alarm off*). A score of 0-points was given for responses that referred to irrelevant facts, fabrications or simple reiteration of the story (e.g. for a ToM story - *Cos it might be there. Can you tell me any more about that? But it ain't there, it ain't there, it's near his bed*; for a non-ToM story - *Because really most people in these shops have lots of alarms so that nothing can be burgled and that's why*). Twenty randomly selected scripts were scored by two independent raters, and the Kendall's

tau-b of the inter-rater reliability was .78 for the physical stories and .80 for the ToM stories, indicating a good to excellent level of agreement.

Cartoons (Happe, Brownell, & Winner, 1999). This task was included as a ToM measure with no verbal structure and minimal verbal input. Six cartoons (with no text) from popular magazines were used. See Appendix VI for examples (which the answers below refer to). Three (the 'ToM' cartoons) involved mental state attributions and three (the 'non-ToM' cartoons) involved physical anomalies and were the control condition. The cartoons selected were matched in difficulty using data collected by the author from a sample of 52 TD eight-year-olds (Fisher et al., submitted). The cartoons were presented to the participant one by one, and they were asked 'Why is this funny?' The cartoon was then left in front of them as they answered. Non-specific prompts (e.g. Can you tell me any more about that?) were used if necessary.

Children's answers were coded on a 0-3 scale. A score of 3-points indicated a full answer, with explicit reference to the target part of the cartoon (e.g. for a ToM cartoon *'Because the dog's barking into here and the man thinks the hose is talking'*; for a non-ToM cartoon *'Because of Minnie Mouse looks at the toilet and it's got the ears that she's got on the toilet'*). A score of 2-points was awarded when the child referred implicitly to the relevant features of the cartoon without making the connections between areas clear (e.g., for a ToM cartoon *'Cos the dog's speaking through the thing and it's coming all the way through and it's coming out of the end bit.'* And what's funny here? *'He can hear the dog'*; for a non-ToM cartoon – *'Cos it looks like it's got ears on the toilet seat'*.) One point was awarded when the child referred to relevant areas of the cartoon without referring, even implicitly, to the anomaly or mental state which makes the cartoon funny (e.g., for a ToM cartoon *'Because the dog's shouting through the hosepipe and it's going all the way through when he's not even barking in the hole'*; for a non-ToM cartoon *'Because it's got two of them bits'*) A score of 0-points indicated reference to irrelevant detail, simple description of the cartoon or 'do not know', (e.g. for a ToM cartoon *'Cos there's the big tree and the man's got glasses and he's looking up and the dog's going 'rurr, rurr''*; for a non-ToM cartoon *'Cos is there something strange in the loo and she doesn't like it?'*). A 3-point coding system was developed because of the range in children's answers and because it was felt to be important to distinguish between those children who clearly understood the point of the cartoon, and those who, whilst



they implicitly referred to the target section, did not make explicit the crucial features of the cartoon.

Twenty randomly selected scripts (49%) were scored by two independent raters, Kendall's tau-b was .86 for the physical cartoons and .75 for the ToM cartoons, indicating a good to excellent level of agreement.

Unexpected Stories (based on Harris & Martin, unpublished): This task consisted of eight short stories, written by the author after the six stories written by Harris. Each of the stories involved a situation that changed (for example a girl who was planning to go to the cinema but then had to go to the dentist instead). In 4 of the stories the protagonist knew about the altered situation, in the other four they remained ignorant, (e.g. a girl went to the doctor thinking she would have an injection, and no one informed her that in fact she would be given the immunisation on a sugar lump). The feelings of the protagonist were described in a simple sentence, either 'She/he felt happy' or 'She/he felt sad'. All the stories are given in Appendix VI. In half of the tasks the emotion was appropriately matched to the situation as seen by the protagonist, in the other half the emotion was inappropriate given the knowledge state of the protagonist. There were therefore four categories of story:

1. **Expected-appropriate (EA)** where the story protagonist knew all the events in the story and the emotion term used was appropriate, given the situation and the protagonist's knowledge of the situation.
2. **Expected-inappropriate (EI)**, which has the same story set up except that the emotion term used was inappropriate (e.g. Jane might feel happy because she was going to the dentist's instead of the cinema).
3. **Unexpected-appropriate (UA)** where the protagonist did not know about a change to events (e.g. Anita did not know that in fact she would be given a sugar lump rather than an injection) and the emotion term used was appropriate given the protagonist's lack of knowledge (so in this case, sad).
4. **Unexpected-inappropriate (UI)** which was the same type of story except that the emotion term is inappropriate, given the protagonist's lack of knowledge. In the Unexpected version of the stories, the appropriateness of the emotion given the protagonist's knowledge state is opposite to the appropriateness of the emotion given the situation.

To control for any differences in story complexity, two versions of this task were used, with the appropriateness or not of the emotions reversed for each story.

Half of the participants received each version. Stories were presented in a randomised order. Initial analyses found no difference between the times for the two story sets, and so data from the two sets were collapsed.

The stories were presented sentence by sentence on a computer, in yellow on a blue background, font Ariel size 24. They were programmed in SuperLab 2.0. Participants clicked on a mouse to move to the next sentence. In between each sentence was a barely discernable break of 150 ms to try and prevent participants inadvertently double clicking over a sentence. A practice story was included which the participants read aloud to the experimenter. If they were unable to read sufficiently well the task was discontinued. This was the case for five members of the MLD group. A simple question about each story was asked after its completion in order to provide participants with motivation to attend to it. Reading times were recorded for each sentence. Due to large individual differences in reading time, a baseline mean reading time was calculated for all participants. This mean reading time per sentence was calculated from 52 neutral sentences that were read by all the participants, and which did not include sentences which changed the situation in the story (in case these sentences were longer for some participants to read due to the change in the direction and meaning of the story). The time to read the final target sentences (i.e. she/he felt happy/sad) was then expressed as a proportion of this mean reading time. This was done in order to minimise differences due to variation in reading speed.

Picture Sequences: This task used picture sequences designed and used by Langdon and Coltheart (1999), after Baron-Cohen, Leslie and Frith (1986). There were 8 picture sequences comprising 4 FB stories (involving scenarios such as sweets falling out of a bag without a boy noticing) and 4 mechanical stories (involving scenarios such as a door slamming, shaking a table and causing a glass to fall off). Stories were depicted in 4-card picture sequences using a simple black and white cartoon style. A set of pictures was used as a practice story, with no interference task. During the articulatory suppression or verbal interference (VI) condition participants were asked to repeat 'baa' and to hold onto a soft toy with their non-dominant hand (in order to control for the effects of having to tap with one hand during the control condition) whilst ordering four of the sequences (two FB, two mechanical), and during the control condition were asked to tap softly on the table with two fingers whilst ordering the picture sequences (two FB, two mechanical). The picture sequences for

which participants tapped or baaed were counter-balanced, and were matched for difficulty after piloting on 10 normal adults. A position score was calculated for each sequence by awarding 2 points if the first card was positioned correctly, 2 points if the last card was positioned correctly and 1 point for each of the second and third cards being positioned correctly. Position scores per sequence ranged from 0-6 for each participant. A run score was also calculated, to reflect the number of pictures that the participants had placed in the correct order, regardless of their positioning within the sequence. This was included in order to compensate for the fact that if a participant ordered 3 out of 4 pictures correctly, but then placed the first or last picture at the wrong end, they would score nothing on the 'position' score. Participants were given a score for their longest correct run (4, 3 or 2). Therefore a participant who had the sequence perfectly right would score 6 on the position score and 4 on the run score. However, it was possible for a participant to score 0 on the position score and 3 on the run score, if three of the cards were in the correct order but were out of position by one. The run score therefore could equal 0, 2, 3 or 4 for each sequence. These scores were summed to give a total score for each picture sequence ranging from 0 – 10.

## **Results**

Due to the inequality in language ability between the two groups, all comparisons were repeated with the VMA matched subgroups. The findings from these are reported after each whole group comparison. Since these subgroup comparisons were inevitably linked with a loss of power, effect sizes will be reported for these VMA-matched group comparisons where appropriate. Findings are reported as pertaining to the motivating research questions, outlined in the introduction. The tasks used, and the reason for their inclusion, are summarised in table 8.3.

**Are tasks that allow access to learnt rules easier for individuals with ASD (as compared to those which cannot be solved by rules)?**

**Eyes and Eyes vocabulary task**

The prediction for these tasks was that the participants with ASD would be relatively unimpaired on the Eyes vocabulary test (which allows the use of learnt rules), whilst still showing deficits on the Eyes task. Table 8.4 gives the means and standard deviations for the ASD and MLD groups on the Eyes task and the Eyes vocabulary test. One child (who had ASD) refused to attempt the Eyes task.

***Table 8.4 Performance on the Eyes task and the Eyes vocabulary test in the ASD and MLD groups; mean (s.d)***

	ASD (N=19)	MLD (N = 22)
Eyes task (max = 28)	16.72 (2.74) <sup>a</sup>	15.86 (3.69)
Eyes vocab task (max = 32)	24.74 (5.33)	24.64 (2.97)

<sup>a</sup>N = 18

There were no group differences on either measure (Eyes vocab  $F(1,39) = 0.006$ ,  $p = .94$ ), Eyes  $F(1,38) = 7.30$ ,  $p = .42$ ). When the analysis is repeated with the verbal ability matched groups, there were still no significant differences. (Eyes task; ASD mean = 16.70 (s.d. = 2.63); MLD mean = 16.76 (s.d. = 2.63); Eyes vocab task; ASD mean = 24.55 (4.84), MLD mean = 25.18 (2.96)). The effect size for the Eyes task was 0.02, and for the Eyes vocab test it was 0.16. These effect sizes were both very low, below that which Cohen (1969) suggests should be considered small effects. It seemed important to identify whether both groups were performing badly on this task, or whether the ASD group was performing particularly well. Baron-Cohen et al. (in press) give results for a group of 15 males with Asperger's Syndrome (AS), as well as TD children aged from 6-12. The males with AS (whose mean VIQ was 102.4, almost exactly the same as the ASD group in this study) had a mean score of 12.6 on the Eyes task. The ASD group in the present study performed 1.24 standard deviations (from the Baron-Cohen et al. data) better, with mean score of 16.7. The ASD group in the present study had a mean CA of 12.13, and a mean BPVS VMA of 12.16. TD children of a comparable CA (the closest are the 10-12 age group in Baron-Cohen's study) scored on average 20.6 (averaged across males and females). The ASD group are therefore performing 1.63 standard deviations worse

these TD children. The MLD group, in contrast, were best compared to the 8-10 year old average, since their average BPVS VMA is 9.05. The average for this group in Baron-Cohen's study is 17.90, and the MLD group performed less than half of one standard deviation worse than this at 15.90. If the matched subgroups were used (and therefore both groups could be compared to the same normal data), then the comparison is between the ASD group with a mean of 16.70 and the MLD group with a mean of 16.80. The best comparison group (taking into account both BPVS VMA and TROG VMA) are the 8-10 year olds in Baron-Cohen et al's study, who had a mean of 17.9. The ASD and MLD groups are both .29 standard deviations below them. However, these are very small groups (both in the present study and the Baron-Cohen et al. study). If we look at all available data, it seems that the ASD group in the present study are performing between the AS and the TD groups from the Baron-Cohen et al. (in press) study, whilst the MLD group perform comparably with TD children of a similar VMA.

Performance on the Eyes task and the Eyes vocabulary task were correlated in the MLD group only ( $r(22) = .53, p < .05$ ). In the ASD group  $r(18) = -.01, p = .97$ ).

#### Think and Not-think task

The prediction from these tasks was that participants with ASD would be relatively unimpaired on the 'Not-think' task, which could be passed using learnt information, whilst being impaired on the 'Think' task, which requires reflection on own mental states.

Table 8.5 gives the percentage of children in each group who were able to do each of the aspects of the Think task. One child in the ASD group refused to do both the Think and the Not-think task. Chi-square tests showed no significant differences in the percentage who could do each individual item (all  $p$ -values  $> .13$ ). Children were awarded a point for each item that they had completed successfully and these were summed to give a total score out of five. There was no group difference on this 'Think' score (ASD mean = 2.61, MLD mean = 3.05,  $F(1, 38) = 1.20, p = .28$ ). However, the VMA-matched groups had a highly significant difference on the Think task (ASD mean = 2.00, MLD mean = 3.18,  $F(1, 25) = 8.90, p < .01$ ). The effect size for this difference is 1.18, which is very large (Cohen, 1969). Thus, children with autism performed very badly on this task when compared to participants with MLD of a similar verbal ability.

**Table 8.5 Percentage of children in each group able to do each aspect of the Think task**

Measure	ASD (N = 18)	MLD (N = 22)
Stated a like and a dislike	72	91
Recalled the like first	72	91
Said they weren't thinking about what they had for breakfast	72	77
Showed evidence of visualising or recalling their thoughts.	22	23
Used mental states terms other than repetition of 'think' or 'thought'?	37	18

Table 8.6 gives the percentage of each group able to do each section of the Not Think task, in which children were asked not to think for 25 seconds. There were no differences between the groups on the percentage for each of these individual scores (all  $p$ -values  $>.12$ ). When the scores were summed to give an aggregate ranging from 0-5, there was no group difference (ASD mean = 2.89, MLD mean = 2.64,  $F(1,38) = .21$ ,  $p = .65$ ). When the VMA – matched subgroups were compared there was still no significant group difference (ASD mean = 2.20, MLD mean = 2.88,  $F(1,25) = 1.00$ ,  $p = .33$ ). The effect size of the difference was .40, which is a small effect, according to Cohen (1969).

**Table 8.6 Percentage of children in each group able to do each aspect of the Not think task**

Measure	ASD (N=18)	MLD (N=22)
Said they had thoughts and reported some.	50	59
Reported some mental activity (including those who were assessed as passing the previous category)	56	68
Mental strategy reported for stopping thoughts.	67	46
Said it was hard to do.	78	55
Gave general explanation of why it was hard to do.	39	36

This task had a potential problem in that those who were the most suggestible might answer that they did not have any thoughts in the hope of fulfilling the expectations of the experimenter. For this reason, the children were asked during the Think task whether they had thought about what they had eaten for breakfast. Those children who said yes were considered particularly suggestible, and so the analyses were repeated with them excluded. This made no difference to the results.

**Are tasks with a well-defined linguistic structure easier for individuals with ASD (as compared to tasks without a linguistic structure)?**

The prediction for these tasks was that the ASD group would be relatively unimpaired on the Strange Stories and 2<sup>nd</sup> order FB (which have a clear linguistic structure), whilst still demonstrating deficits on the ToM cartoons.

**ToM cartoons and stories**

Table 8.7 gives the mean scores for the ToM and control Physical cartoons and the ToM and control Physical stories. One child in the MLD group refused to do the Stories task. A MANOVA looking for group x condition interaction was performed for the Cartoons task. In the whole group there were no main effects of group on either section of the Cartoons (ToM cartoons  $F(1,39) = .35, p = .56$ , Physical cartoons  $F(1,39) = 2.31, p = .14$ ). There was also no significant interaction ( $p = .12$ ), although inspection of the data reveals that there was a tendency towards an interaction, with the ASD group doing better on the Physical cartoons than the ToM, and the MLD group showing the opposite pattern.

***Table 8.7 Mean scores on the Cartoons and Stories for the ASD and MLD groups; mean (s.d)***

	ASD (N = 19)	MLD (N=22)
ToM cartoons (max = 9)	5.21 (2.34)	6.05 (2.07)
Physical cartoons (max = 9)	5.64 (2.24)	4.86 (2.82)
ToM stories <sup>a</sup> (max = 6)	3.05 (2.58)	1.76 (1.34)
Physical stories <sup>a</sup> (max = 6)	3.94 (1.58)	2.71 (1.38)

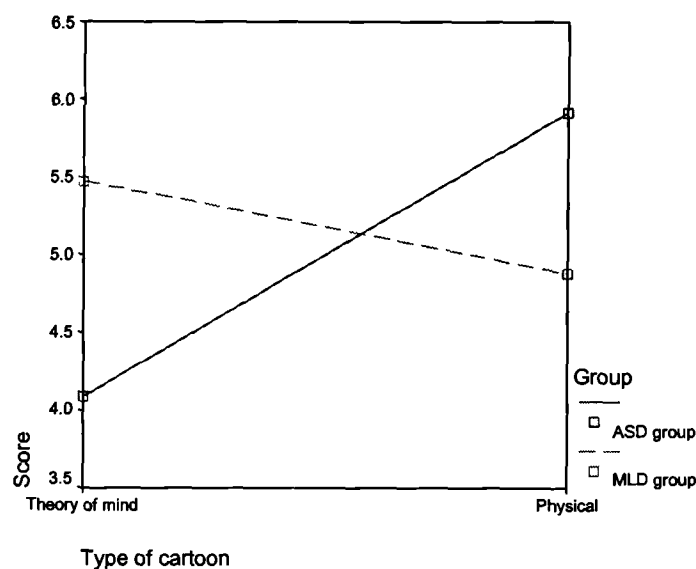
<sup>a</sup>N = 21 for MLD group

With the VMA-matched subgroups, there were still no group differences on the two sections of the task, although there was a moderate effect size (0.64) for the ToM cartoons and a small effect size (0.42) between the Physical cartoons. There was, however, a significant interaction between cartoon type (ToM versus physical) and group ( $F(1,26) = 4.27, p < .05$ ). Figure 8.1 illustrates this interaction.

On the Strange Stories, a MANOVA looking for a group x condition interaction found a main effect of group with the ASD group performed significantly better than the MLD group in both conditions (ToM stories  $F(1,38) = 7.82, p < .01$ , Physical stories  $F(1,38) = 6.92, p < .05$ ). There was no significant interaction between group and story type. When the VMA-matched subgroups were compared, there were

no significant group differences on either condition, effect sizes were 0.51 for the ToM stories (a small to moderate effect) with the ASD group performing slightly better than the MLD group, and 0.02 for the Physical stories. There was no significant interaction between group and story type.

**Figure 8.1** *Illustration of the interaction between cartoon type and group for VMA-matched subgroups*



## 2<sup>nd</sup> order FB

Initially the two 2<sup>nd</sup> order FB tasks were analysed separately. Participants were given 1 point for answering the FB question correctly and an additional point if they then justified their answer correctly. There was no significant difference on either task (both p-values >.37). Table 8.8 gives the percentage of each group who answered each question and justification, as well as the percentage who failed a control question and were therefore excluded from that task.

**Table 8.8** *Percentage of each group who passed each FB task, passed the justification question or failed the control question.*

	ASD (N=19)	MLD (N=22)
Chocolate story passed FB	94.8%	81.8%
Chocolate story passed justification	78.9%	77.3%
Chocolate story failed controls	5.2%	13.6%
Circus story passed FB	63.2%	86.4%
Circus story passed justification	57.9%	63.6%
Circus story failed controls	21.1%	13.6%



There were no significant differences between the groups. The scores were then combined to create a FB aggregate score, in which the participants were given one point for each FB question and one point for each justification question that they answered correctly. The FB score therefore ranged from 0-4. Initially those who failed a control question were credited with their score for only the tasks on which they passed the control. There was no significant group difference on this score (ASD mean = 2.95 (s.d = 1.26), MLD mean = 3.09 (s.d. =1.23)  $F(1,39) = 14, p = .72$ ). This was still the case if those who had failed control questions for one task were credited with double their score for the task on which they had passed controls questions.

When the VMA matched groups were used, 2/11 of the ASD group and 8/17 of the MLD group passed all the FB questions and justifications. This difference did not reach significance (Fisher's Exact test  $p = .23$ ). However, the effect size of this difference is very large (1.64) indicating that had the sample size been larger this magnitude of difference would have reached significance.

### **Is there evidence that individuals with ASD are using post hoc strategies in mental state attribution?**

#### **Unexpected Stories**

The prediction for this task was that the children with ASD would make mental state attributions only when these were clearly required, and hence do so in a post hoc fashion. This would be evident in their reading times (as a measure of processing speed), which would be longer when based on a misinformed protagonist's mental state as compared to an informed protagonist's mental state, since the children with ASD would not be keeping track of the protagonist's mental states online. Five children with MLD were not able to read well enough for this task to be carried out.

Initial inspection revealed that reading times differed significantly between the two experimental groups (with the ASD group reading faster than the MLD group), but also that the standard deviations for these reading times were very large, and different in the two groups. Therefore, a baseline mean reading time per sentence score was calculated for each child from the 52 'standard' sentences that formed the structure to all of the stories. The ASD group were significantly faster readers (ASD mean = 4105 secs, MLD mean = 6941 secs,  $F(1,34) = 26.54, p < .001$ ). The reading times for the target sentences (those containing the emotion term) were then expressed

as a proportion of this total reading time. All reading times less than or equal to 200 ms were excluded, as they were assumed to be due to a participant skipping a sentence, or over-enthusiastically clicking the mouse (2 occurrences out of 988 sentences in the ASD group, and 17 occurrences out of 936 sentences in the MLD group). Table 8.9 gives the mean times (expressed as a proportion of the mean reading time for the 52 ‘standard’ sentences) for each story condition. There were two stories in each condition and thus each mean represents the average of the two stories. If a participant’s time for one story was excluded due to being less than 200 ms, their time for the other story of the same condition was used in place of the mean.

**Table 8.9 Mean reading times for each story condition as a proportion of total reading time: mean (s.d)**

Story condition	ASD (N=19)	MLD (N=17)
Expected-appropriate (EA)	.43* (.16)	.31* (.05)
Expected-inappropriate (EI)	.61 (.55)	.37 (.08)
Unexpected-appropriate (UA)	.49* (.23)	.30* (.27)
Unexpected-inappropriate (UI)	.50** (.27)	.34** (.10)
Group differences * $p < .05$ , ** $p < .01$		

Several of these scores violated the normality assumption as tested by the Kolmogorov-Smirnov test. In addition, the assumption of homogeneity of variance is violated in several cases. Therefore, non-parametric tests were used. An initial Mann-Whitney test found that the groups differed significantly, (with the ASD group taking longer to read the sentences) on all of the conditions apart from EI. With the VMA-matched groups there were no significant differences on any of the conditions (all  $p$ -values  $> .14$ ).

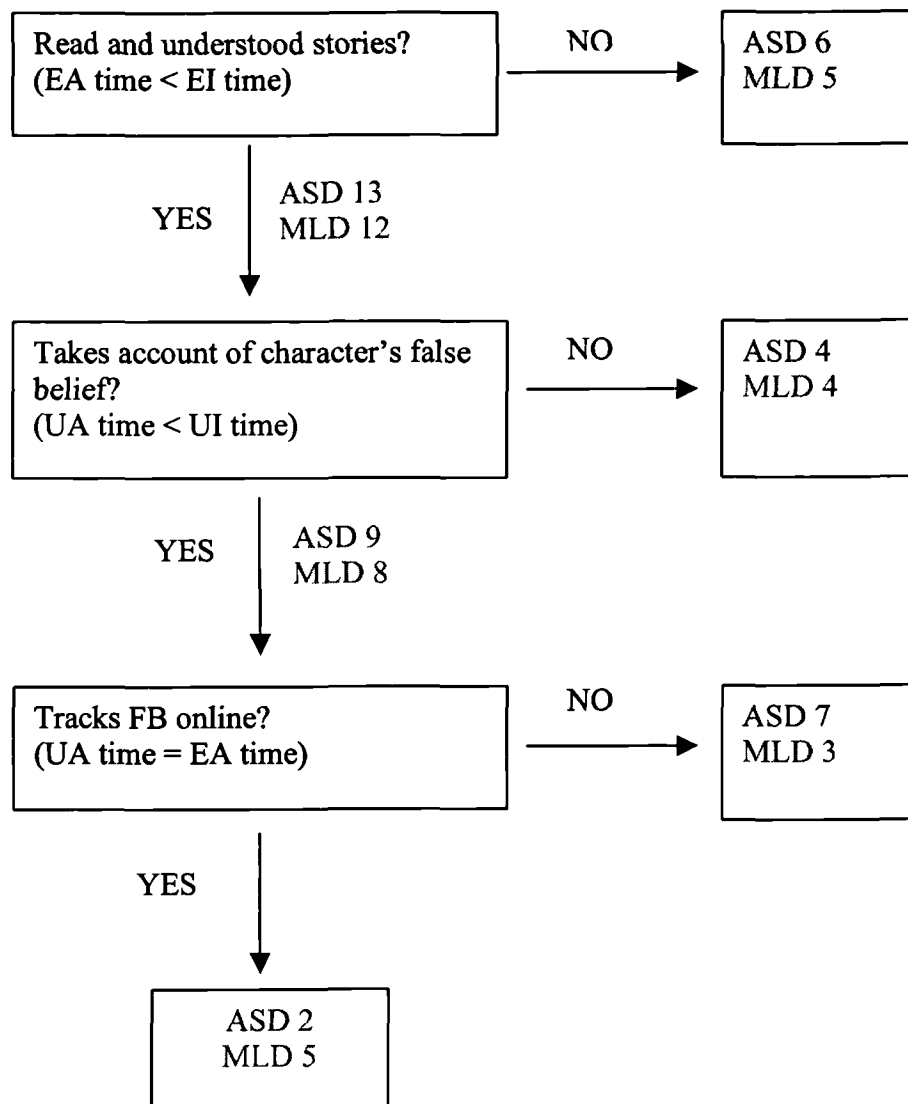
Failing to perform as expected on certain features of this task make later comparisons meaningless – for example, if there is no evidence that a participant is attending to and understanding the stories, then a difference in reading time between conditions is likely to be spurious. Participants can be divided into four groups. Those who did not read the sentence referring to the appropriate emotion faster than the inappropriate emotion in the Expected condition make up a group who did not appear to be taking the emotional content of the story into account; we therefore have no evidence that they were engaging with and understanding the story. Of those remaining, those who read the inappropriate emotion faster than the appropriate emotion in the Unexpected condition did not appear to be taking the protagonist’s FB

into consideration. In the final group (who appeared to be both understanding the story and taking account of the FB) the relative speed of reading the appropriate emotion sentence in the Expected or Unexpected conditions tells us, potentially, whether participants are attributing mental states as the story progresses, or are only calculating the characters' thoughts when emotional information that conflicts with reality occurs. If they were building up the mental model as the story progresses, the emotion sentence should not come as a surprise in either story type, and reading time for the emotions sentences should not differ. If, however, participants are calculating mental states in a 'post hoc' fashion, then it should take them longer to read the emotion sentence in the 'Unexpected appropriate' condition than in the 'Expected appropriate' condition. Figure 8.2 gives the numbers who fit into each of these groups in the ASD and MLD groups.

As can be seen from figure 8.2, similar numbers of those with ASD and those with MLD do not appear to be understanding the story properly (6 ASD, 5 MLD). Four children in each group do not appear to be taking account of the FB. Of the remaining children, 7/9 of the ASD group are slower to read the appropriate emotion in the Unexpected condition, indicating that they may be using a post hoc strategy, whilst only 3/8 of the MLD group show the same pattern. This difference in numbers was not significant. However, if we compare the group mean reading times for just those remaining children (who are the only ones for whom the comparison is really valid), the time to read the appropriate emotion sentences in the Unexpected and Expected conditions are significantly different in the ASD group ( $Z = -2.07$ ,  $p < .05$ ) and not in the MLD group ( $Z = -0.98$ ,  $p = .33$ ).

Since the unexpected condition is effectively a FB task, those who read the appropriate term slower than the inappropriate term are effectively failing an implicit FB task (as they are treating the emotion term as appropriate for the situations rather than for the protagonist's perspective). All four individuals in the ASD group who read the unexpected appropriate term slower than the unexpected inappropriate term passed both questions and both justifications of the 2<sup>nd</sup> order FB tasks. None of the 4 individuals in the MLD group did.

**Figure 8.2 Performance on the Unexpected Stories task**



**Is there evidence that any online strategies are verbally-mediated?**

The Picture Sequences task used articulatory suppression with the aim of interfering with any verbal strategies that the ASD group might be using to solve ToM tasks. The prediction was, therefore, that the ASD group would be selectively impaired on the FB scenarios (as compared to mechanical scenarios) in the verbal interference (VI) condition, as compared to a control tapping condition.

**Picture sequences**

Two children with ASD and one child with MLD did not complete this task. The picture sequences were presented in pairs of FB or Mechanical sequences matched in difficulty after piloting on 10 normal adults. There were two pairs of FB

and two pairs of Mechanical sequences. The condition (i.e. VI or tapping) for each picture sequence was counterbalanced across participants, as was the order that they were presented in. There were no differences according to which condition or order the pictures were presented in (all p-values >.39), and therefore the data were collapsed so there were four categories, a FB VI condition (in which the participant had to repeat the word baa), a FB control condition (in which the participant had to tap on the table), a Mechanical VI condition and a Mechanical control condition. Each picture sequence was scored out of 10. These scores, and effect sizes comparing performance on the two FB and Mechanical conditions within each group are given in table 8.10. There were no group differences on any of these scores (all p-values >.11). A repeated-measures ANOVA found no differences within groups between the VI or control conditions of each task, or between the FB and Mechanical picture sequences. When the VMA-matched groups were used there were still no group differences (all p-values > .37).

***Table 8.10 Scores for the Picture Sequences task for the ASD and MLD groups: mean (s.d)***

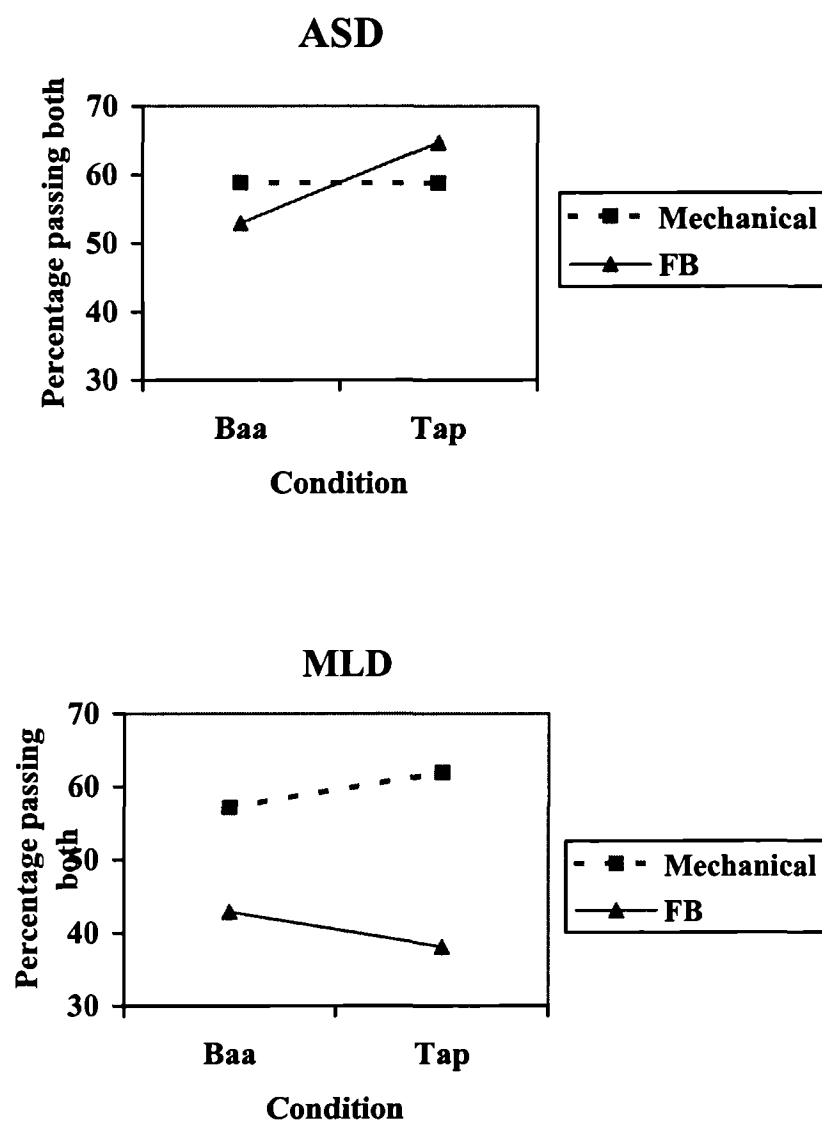
	ASD (N = 17)	MLD (N = 21)
FB verbal interference (max = 10)	7.76 (2.70)	7.02 (3.08)
FB control (max = 10)	8.50 (2.34)	7.21 (2.56)
Mechanical verbal interference (max = 10)	8.03 (2.86)	8.00 (2.71)
Mechanical control (max = 10)	8.26 (2.30)	8.14 (2.55)
Effect size- FB	0.29	0.07
Effect size- Mechanical	0.09	0.05

It may also be informative to compare the number of individuals in each group who were able to complete both sequences correctly within a particular condition. Figure 8.3 shows the percentage of participants who ordered both sequences correctly in each condition.

There were no significant group differences between the numbers who passed both or those who failed one in any of the conditions, but there was a marginal trend in the FB control condition ( $\chi^2 (1) = 2.66, p=.10$ ), with the ASD group performing better than the MLD group. Again there was a small effect size between the FB conditions in the ASD group (0.23), and no effect in the MLD group (0.09 for both FB and Mechanical) or between the Mechanical condition in the ASD group (0.00).

Figure 8.3 illustrates the comparison between conditions – verbal interference (baa) and control (tap) in the two different groups.

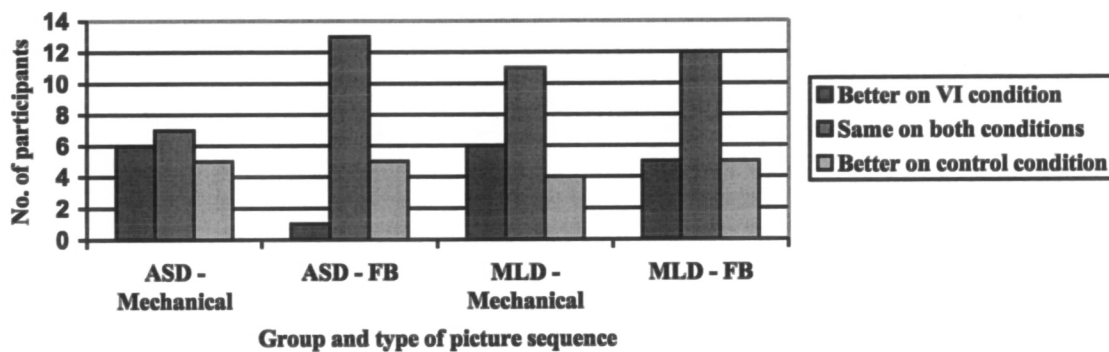
*Figure 8.3. Percentage passing both sequences correctly in each condition, split by group*



Again there is evidence that there is a small difference between the verbal interference and control condition in the ASD group only, but that the sample size of this study is too small for this to be significant. In order to look more closely at individual performance, the number of sequences completed correctly in the verbal interference condition was subtracted from the number completed corrected in the control condition to give a difference score, indicating whether the individual

performed better on the verbal interference or control condition. Figure 8.4 illustrates the number of participants in each group who performed better in the VI condition or control condition, or the same in both conditions. Only a single participant in the ASD group performed better on the VI condition on the FB sequences, whilst 6 showed this pattern on the mechanical sequences. This difference is not significant.

**Figure 8.4. No. of participants who performed better on each condition in the ASD and MLD group.**



#### **How does performance on the ToM tasks relate to ability?**

In line with previous work looking at the relationship between ToM and language in ASD (Happé, 1995; Tager-Flusberg & Sullivan, 1994 and Study 1, this thesis) a prediction of this study was that performance on the ToM tasks would be related to verbal ability (particularly in the ASD group). If this reflects task-specific factors only, the ToM tasks with high language demands (e.g. the Strange Stories) should be more strongly related to measures of verbal ability than those with lower language demands. If language is developmentally linked to performance on these ToM tasks, then performance on less verbal ToM tasks (e.g. the ToM cartoons and Eyes task) may be just as strongly linked to language as the more verbally based tasks.

Correlations between the individual tasks and ability measures were not easily interpretable due to the large number of tasks and therefore the inflated risk of a Type 1 error, as well as the small sample size. To examine the relationship between tasks and with ability, four groups of tasks were created, on the basis of the theoretical questions motivating this study. These were: ToM tasks with a strong language structure (ToM stories, 2<sup>nd</sup> order FB); ToM tasks which are soluble by rules (Eyes vocab, Not-Think); ToM tasks without a strong language component (ToM cartoons,

Eyes); Non-ToM control tasks (Physical cartoons and stories). The range for each task varied greatly, and so, for comparability, the scores were divided up by quartiles for the whole group. Those in the top 25% for each task were awarded 3 points, those in the 50<sup>th</sup>-75<sup>th</sup> percentile were awarded 2, those in the 25<sup>th</sup> to 50<sup>th</sup> percentile were awarded 1, and those below the 25<sup>th</sup> percentile were given 0. The scores for the two tasks in each set were then summed to give a score ranging from 0-6. This score therefore gave a measure of how well a particular participant had done as compared to all the other participants in this study.

These scores were correlated with the age, ability and working memory measures. These values are given in table 8.11. The VMA scores alone were used for the TROG and BPVS to ease comparison.

**Table 8.11 Correlations between task aggregates and ability measures in the ASD and MLD groups (ASD in bold)**

	ToM with language structure		ToM – no/minimal language input		ToM – rule based		Non-ToM controls	
Diagnostic group	<b>ASD</b> (N=19)	MLD (N=22)	<b>ASD</b> (N=19)	MLD (N=22)	<b>ASD</b> (N=19)	MLD (N=22)	<b>ASD</b> (N=19)	MLD (N=22)
CA (years)	<b>.48*</b>	-.12	<b>.42</b>	<b>.56**</b>	<b>.33</b>	.40	<b>.39</b>	.09
Raven's CPM (raw score)	<b>.32</b>	-.06	<b>.62**</b>	.08	<b>.21</b>	<b>.49*</b>	<b>.56*</b>	.14
BPVS VMA (years)	<b>.54*</b>	-.13	<b>.49*</b>	.38	<b>.40</b>	.21	<b>.48*</b>	.05
TROG VMA (years)	<b>.52*</b>	.17	<b>.64**</b>	.21	<b>.69**</b>	<b>.43*</b>	<b>.06</b>	.06
Forwards digit span	<b>.27</b>	-.19	<b>.11</b>	.17	<b>.06</b>	<b>.48*</b>	<b>.49*</b>	.23
Backward s digit span	<b>.50*</b>	.09	<b>.53*</b>	.01	<b>.47*</b>	.29	<b>.40</b>	.33

\*p <.05, \*\*p<.01

Language and ability were strongly related to the rule-based ToM tasks in both the ASD and the MLD groups. However, in the case of those ToM tasks with a strong language structure, the relationship was much more evident in the ASD than in the MLD group. In the ToM tasks without language input, non-verbal ability, verbal ability and working memory were related to performance in the ASD groups, but not in the MLD group (where only CA was significantly correlated with this variable), although BPVS VMA had a moderate effect size. On the Non-ToM control tasks,



language and ability were again more strongly related in the ASD group, with the notable exception of the TROG.

Overall, the relationships with ability are more consistent and of larger magnitude in the ASD group than the MLD group, with the exception of the rule-based tasks. Working memory (as measured by the backwards digit span) appears to be much more important in the ASD group than in the MLD group, particularly on the language and non-language based ToM tasks. The TROG showed a particular pattern of associations in the ASD group, strongly associating with all the different ToM task aggregates, but not at all with performance on the non-ToM control tasks. Fisher's z-transformation found significant differences between the correlations between the TROG and ToM-rule based tasks, and the TROG and the non-ToM control tasks ( $z = 2.32, p < .05$ ) and the ToM-no language input v. the non-ToM control tasks ( $z = 2.06, p < .05$ ). The difference between the correlations between the TROG and the language based ToM tasks and the TROG and the non-ToM control tasks was not significant ( $z = 1.53, p > .10$ ). The TROG was the only task to show such a strong difference between the ToM groups and the non-ToM controls within a group, in all other cases the magnitude of the correlations were not greatly different. It is striking that language, ability and working memory are so strongly related to the no-language ToM tasks in the ASD group – there are tasks (the Cartoons and the Eyes) in which the language input is minimal and it is hard to see how verbal strategies may be used.

#### Relationships between ToM and Non-ToM control tasks

Again, to minimise the chance of the Type 1 error, the same aggregates of tasks were used to look at the relationships between different types of tasks. Table 8.12 gives the correlations between the different sets of tasks, by group.

**Table 8.12 Correlations between groups of ToM tasks and the non-ToM control tasks. ASD (N=19) above the diagonal and in bold, MLD (N=22) below diagonal.**

	ToM with a language structure	ToM – rule based	ToM – no language	Physical controls
ToM with a language structure		<b>.63**</b>	<b>.42</b>	<b>.44</b>
ToM – rule based	-.04		<b>.51*</b>	<b>.13</b>
ToM – no language	.14	.21		<b>.25</b>
Physical controls	.20	.02	.37	

\* $p < .05$ , \*\* $p < .01$

In the ASD group there are strong relationships between the three sets of ToM tasks – although the relationship between the ToM–no language set and the ToM–language set is not significant, it is of a moderate effect size. In contrast, the non-ToM control tasks related (non-significantly again, but with a moderate effect size) only to the language-structured ToM tasks.

In the MLD group the pattern is quite different, there are no differences in the relationship between the ToM tasks and the non-ToM control tasks. The correlations between the ToM-language tasks and the ToM-rule based tasks are significantly different in the two groups ( $z = 2.30$ ,  $p < .05$ ). The differences between the other correlations do not reach significance (all  $z$ - values  $< 1.06$ ).

## **Discussion**

I will address each of the key theoretical questions posed in the introduction in turn.

### **Are ToM tasks that allow access to learnt rules or information easier for participants with ASD?**

A prediction of this study was that, if able individuals with ASD are using their language skills to accumulate a set of learned rules for social situations, enabling them develop some understanding of mental states, then tasks that potentially allow access to learnt information (such as asking for definitions of mental states) should be easier for them. Tasks in which learnt rules are insufficient (e.g. the Eyes task and the ‘Think’ task) should be harder. The prediction was partially supported. There were no differences between the ASD and MLD groups (whether the whole sample was used or just the VMA-matched subgroups) in the Eyes vocabulary task and the Not-think task, both of which could potentially be solved by learnt rules. There were group differences on the Think task, but, unexpectedly, not on the Eyes task itself. However, comparisons with data for TD children of a similar VMA (from Baron-Cohen et al, in press) suggested that the ASD group were performing worse than the MLD group, However, the ASD group are also performing better than the group with Asperger’s Syndrome from the Baron-Cohen et al. (in press) study, indicating that they are less impaired than might have been expected from previous studies. This may be because the present sample was pre-selected for their (relatively) good ToM ability, having all passed 1<sup>st</sup> order FB.

The other paired tasks on which there was a difference on one task but not on the other were the Think and Not-think tasks. We might have expected the Think task to be easier than the Not-Think task, since in this task there is no conflict between what the participant is being asked to do and what it is possible for them to do. Certainly in Flavell's original work (Flavell et al., 2000) the five- and eight-year-olds had more success on the Think task than on the Not-Think task. The MLD group in this study showed a similar pattern. However this was not the case in the ASD group – in fact scores on the Not-think task were slightly higher than those on the Think task. Given that the results of the Think task indicated that participants in the ASD group did have difficulty reflecting on the content of their thoughts, it seems odd that they were not more disadvantaged in the Not-think task, unless we assume that they were able to use learnt rules with this task, but not with the Think task. It is possible to see how a child may have learnt that it is not possible not to think, and therefore would report back that they were unable to stop thinking, without actually reflecting on their mental states. To date, there has been very little research on the ability of individuals with ASD to 'read their *own* minds', and this data provides some evidence that it may be an area in which deficits remain even in those who are able to develop some understanding of other minds.

### **Are tasks with a verbal structure easier for individuals with ASD?**

As predicted, on tasks which had a well-defined verbal structure (Strange Stories and FB), the participants with ASD were unimpaired (or, in the case of the Strange Stories, much better) when compared to the MLD group. By contrast, on a task with no verbal scaffolding (the Cartoons), the VMA-matched subgroups showed an interaction between task and group, with the ASD group finding the ToM cartoons harder than the Physical cartoons, and the MLD group showing the opposite pattern.

Thus the present group of participants appeared to be relatively unimpaired on ToM tasks (as compared to non-ToM control tasks) with a verbal structure, whilst still showing impairment on non-verbal tasks. This is consistent with the idea that individuals with ASD are using a verbal strategy, application of which is greatly eased by presentation of a verbally defined task. However, it is also possible that the verbal tasks are easier for a different reason – for example, individuals with ASD are known to have a tendency towards weak central coherence (Shah & Frith, 1993), which might make it hard for them to bring together the elements in a cartoon picture and

extract meaning from it. However, this would not explain why the participants with ASD were able to perform well on the Physical cartoons, which presumably make the same demands on central coherence as the ToM cartoons. Alternatively, it is possible that individuals with ASD were not using a verbal strategy spontaneously, but were able to make use of a verbal structure when one is provided.

### **Are individuals with ASD using a post hoc strategy for attributing mental states?**

Evidence from the Unexpected Stories task indicates that some individuals with ASD may be using a post hoc strategy for attributing mental states, rather than taking the perspective of the protagonist throughout a story as it appears that normal adults do (Harris & Martin, unpublished). This task did not ask participants any questions about mental states, all analyses were based on the relative reading times for sentences with emotional content. It therefore avoided cuing the participants by asking a leading question, and did not draw their attention towards the mental states as a key part of the story. This task aimed to investigate the strategies that children with ASD use spontaneously, when they are not explicitly asked to make any deductions or predictions about mental states. The participants split into 4 groups in this task. About a quarter of both groups did not appear to be reading the stories for meaning (insofar as reading times did not reflect congruency effects of emotion and story), and they were therefore disqualified from further analysis. Four individuals in each group appeared to be ignoring the protagonist's lack of knowledge, and instead judged emotions to be appropriate or not by the emotional content of the situation. These individuals could be said to be failing implicit first order FB tasks. Of course, all participants were chosen for this study because they passed standard FB tasks, and therefore it is quite striking that some did not demonstrate this knowledge in a more implicit task. In addition, all four of the individuals in the ASD group who failed to demonstrate integration of the mistaken belief in the Emotional stories passed both 2<sup>nd</sup> order FB tasks and their justification questions, whilst all four of the individuals in the MLD group failed at least one of the 2<sup>nd</sup> order FB questions or justifications. Those individuals with ASD would appear to be a group who, whilst they can make accurate mental state attributions in explicit tasks such as these, do not do so spontaneously or without prompting, even in fairly simple verbal tasks. This may indicate that the ASD group are using a different route to passing the tasks than are the MLD group, one

which comes into play when they are asked a question but not when they are not explicitly asked about mental states.

This may indicate that some individuals with ASD are slowed by the need to make sense of emotional information that matches the protagonist's (false) belief, rather than reality. This contrasts with young children (Rall & Harris, 2000) and adults (Harris & Martin, unpublished), who appeared to integrate mental states online, as did most of the individuals with MLD in this study. It appears that ToM in most people (including those with MLD) may be something that almost seems to be obligatory, occurring without the need for effort or intention, whilst in individuals with ASD it may be a cognitive, intentional and effortful process.

### **Are individuals with ASD using a verbal strategy online?**

Some researchers have suggested that individuals with ASD are solving FB tasks using a verbal online strategy (Tager-Flusberg, 2002, Happé, 1995), for example talking themselves through a task. The Picture Sequences task therefore included an articulatory suppression condition, designed to interfere with verbal strategies, to see this selectively impaired performance on a non-verbal FB task. There was some preliminary evidence for the effectiveness of this interference in the present study. The verbal interference task selectively impaired performance (as compared to a 'tapping' control condition) on FB scenarios in the ASD group only.

However, this difference was not significant, and as such firm conclusions cannot be drawn. Looking more at individual performance, there was only a single individual with ASD who performed better on the verbal interference condition than the control condition of the FB sequences, whereas on the Mechanical sequences (and in the MLD group in both FB and Mechanical conditions) numbers were spread quite equally between those who performed better on each condition. These results need to be repeated with a larger sample size.

### **How do ToM tasks relate to ability and to each other?**

The ToM tasks were divided into 3 types of task to look at the relationship with ability and with each other. The sets were ToM-language (ToM stories and 2<sup>nd</sup> order FB), ToM-rule based (Eyes vocab and Not-Think) and ToM-no language input (ToM cartoons and Eyes). A Non-ToM control aggregate was also formed for comparison (Physical cartoons and Physical stories). The rationale for dividing the

tasks was that if there was a strong relationship between language and the ToM-no language input tasks, this would suggest a developmental role for language in ToM (as compared to a relationship between language and the more language based ToM tasks, which could equally well reflect task factors, or an online strategy).

Non-verbal and verbal ability were strongly related to all four groups of tasks in the ASD group. There was no evidence that this relationship was specific to the language based tasks. In the case of grammatical ability the relationship was specific to ToM tasks, there was no relationship with the physical control tasks. Verbal and non-verbal ability was related much less strongly to task performance in the MLD group, with the exception of the rule-based ToM tasks. This indicates that in the ASD group, the relationship between language and ToM is likely not to be due solely to the language demands of the tasks, or even due to verbal strategies used online (since it is hard to envisage how verbal strategies might help performance on non-verbal tasks such as the Eyes task). Instead, it indicates, in line with the findings from Study 1 (chapter 3), that language – and particularly grammar - is important developmentally in the acquisition of some understanding of mind in ASD.

There was also evidence that working memory is important in ToM task performance in ASD – both the language based and non-language based ToM tasks were significantly correlated with working memory in the ASD group, whilst there was no evidence for a relationship in the MLD group. This may again support the hypothesis that individuals with ASD are using an alternative strategy for ToM, one that makes more cognitive demands upon the individual.

## **Conclusion**

The results of this study indicate that high-functioning children with ASD were unimpaired as compared to a control group of children with MLD on ToM tasks that had a defined linguistic structure or that allowed the application of learnt rules. Differences between VMA and non-verbal ability matched ASD and MLD subgroups remained on tasks which had no external verbal structure or which required the participants to reflect on and report their own recent mental states. However, these differences were only apparent when the VMA-matched subgroups were used, indicating that a high verbal ability can help individuals with ASD compensate for many of their ToM difficulties across all types of task. Of course, there was no TD control group and so it remains a possibility that the MLD group performed badly on

the ToM tasks, rather than that the ASD group performed well. There is no specific reason, however, to assume that this MLD group were particularly impaired in their socio-cognitive skills or understanding – nor that they would be specifically impaired in tasks with a well-defined verbal structure or tasks on which learnt rules could be used. In addition, many of the comparisons looked at the relative performance on different tasks within (rather than between) groups (e.g, ToM stories vs. Physical stories). It therefore seems that, with some qualification, we can conclude that difficulties in advanced ToM persist in this group of high-functioning people with ASD, when compared to verbal age and non-verbal ability matched MLD controls, and that their success on some tasks relies on a strong verbal structure or the potential to use learnt rules. Language was related to performance in all sets of ToM tasks in the ASD group, indicating that language is important developmentally as well as online. This was particularly true of grammatical ability, which was not related at all to performance on the physical control tasks, but which was strongly related to performance on all groups of ToM tasks. However, it is also the case that areas of surprising strength exist in individuals with ASD, and any compensatory strategy they may be using to solve ToM problems is quite effective, and works over an impressive range of tasks.

There was evidence that some individuals with ASD were attributing mental states only as and when required, rather than taking the perspective of a protagonist throughout a story, as TD individuals appear to do (Rall & Harris, 2000). There was some tentative evidence that some individuals may be using verbal strategies during tasks which can be disrupted by concurrent articulatory suppression. These results emphasise the diverse nature of high-functioning individuals with ASD – a range of different strategies are likely to be used. Both of these results would benefit from replication with larger sample sizes.

In conclusion, a group of high-functioning individuals with ASD were remarkably unimpaired on many ToM tasks when compared to controls with MLD. However, group differences remained, particularly on tasks with minimal verbal scaffolding and on which learnt rules were not useful. There was also some evidence that some individuals with ASD were integrating mental states in a post hoc fashion only, or using verbal strategies. Along with studies 1 and 3, this study found that language, and particularly grammar, appears to be very important in enabling individuals with ASD to develop a better understanding of mind. It appears more

likely that this is developmental progression from language to ToM rather than due to either the language demands of the tasks themselves or due to better ToM allowing the development of better language. It may be that high-functioning individuals with ASD have a general tendency *not* to compute or take account of mental states, but *can* do so in many circumstances when and if required, and particularly if supported by verbal scaffolding. This post hoc processing would presumably be more useful in experimental tasks than in real life behaviour, since real life situations do not wait around for answers to be computed, nor necessarily cue such problem solving, but instead require immediate reactions. We would expect, therefore, that even individuals with ASD who pass tests of FB will continue to be impaired on more naturalistic measures of ToM understanding as compared to individuals with learning difficulties. This hypothesis was tested in study 5, reported in the next chapter.



## **CHAPTER 9**

### **STUDY 5: THEORY OF MIND AND EXECUTIVE FUNCTION IN REAL LIFE**

#### **Introduction**

This thesis has so far investigated ToM in children with ASD and MLD, and its relationship to EF and language. This chapter looks at teacher ratings of the real life behaviour of children with ASD and MLD, in relation to performance on experimental tasks.

As discussed in chapters 3 and 8, we know that some individuals with ASD pass ToM tasks, and some individuals with MLD fail these same tasks. We also know that those individuals with ASD who pass basic ToM tasks show experimental evidence of better mentalising ability on other tests. However, we cannot tell from this whether there is a genuine difference in the degree of real life social impairment between those who pass and fail in either group, or whether their superior performance on ToM tasks is simply due either to luck or to a non-ToM problem-solving strategy used by some individuals with ASD. Equally, FB task failure in MLD groups may be due to problems with task-specific demands, and hence not reflect real life functioning.

The relationship between real life social impairment and experimental performance on ToM tasks in ASD has received surprisingly little investigation to date. As discussed in chapter 8, Frith et al. (1994) found a significant group difference on teacher ratings of behaviours that appeared to require a ToM between individuals with ASD who passed ToM and those who failed. They did not find a corresponding difference between individuals with MLD who passed and failed ToM tasks. However, their sample sizes were small, particularly in the MLD group. Hughes, Soares-Boucaud, Hochmann and Frith (1997) found a relationship between teacher and therapist ratings and mentalising abilities in a group of 21 French children with PDD, some of whom had autism. Fombonne, Siddons, Achard et al. (1994) also found a relationship between passing socio-cognitive tasks and performing better on scales of social (and anti-social) behaviour requiring mentalising in a group of French adolescents and adults with autism. This difference became non-significant when the

difference in verbal ability between the passers and failers was accounted for. However, since differences in verbal level appear to be inherent between groups of ToM passers and failers with ASD, controlling for such differences may well be theoretically problematic, as discussed in chapter 8 (see also Miller & Chapman, 2001).

The everyday demonstration of executive functioning in ASD has received only limited investigation to date. Turner (1997) found that repetitive behaviour in ASD was related to performance on EF tasks – specifically to tasks of generativity and perseveration. Turner also found no difference in repetitive behaviour between individuals with ASD who passed ToM tasks as opposed to those who failed, indicating that the difficulties in ToM seen in ASD are unlikely to account for repetitive behaviour. In addition, despite much research (discussed in chapter 4) on the relationship between experimental measures of EF and ToM, to date there has been no investigation of whether measures of everyday ToM functioning relates to everyday EF, as opposed to experimental measures of ToM.

This study therefore had two main questions: Are there differences between individuals with ASD and those with MLD on teacher ratings of everyday behaviour thought to require a ToM, or on behaviour thought to be indicative of good executive functioning? Secondly, how do teacher rating on scales of everyday ToM and EF relate to performance on experimental tasks, specifically FB tasks, and to each other?

## **Method**

**Participants:** Participants had all taken part in either Studies 2 and 3, or Study 4. They therefore all had a VMA as measured by the BPVS of more than 4 years 3 months. All of the MLD group, and all but 6 of the ASD group were recruited from schools for children with special educational needs in the greater London area. An additional 6 children with autism and Asperger's Syndrome were volunteers from a previous study. Out of the ASD group, 24 had a diagnosis of autism or an autistic spectrum disorder, 8 had a diagnosis of Asperger's Syndrome and the remaining 5 (who were from a single school and who were all being educated in classes for children with autistic spectrum disorders) were described on their records as having social and communication disorders. These children were assigned to the ASD group after completion of a checklist of symptoms based on DSM-IV after discussion with their teachers.

Table 9.1 shows the participant characteristics of the two groups who participated in this study. The groups differed significantly on CA ( $F(1,67) = 6.51, p < .05$ ), BPVS standardised score ( $F(1,67) = 9.14, p < .01$ ) and the Raven's Matrices raw score ( $F(1,67) = 8.44, p < .01$ ). There were no differences on the BPVS VMA ( $F(1,67) = 1.91, p = .17$ ), the TROG VMA ( $F(1,67) = .18, p = .68$ ) or the TROG age-adjusted score ( $F(1,67) = 2.51, p = .12$ ).

**Table 9.1 Characteristics of participants in the ASD and MLD groups: mean (s.d.)**

	ASD (N=37)	MLD (N=32)
CA (years)	10.95* (2.35)	12.17* (1.45)
BPVS VMA (years)	8.77 (3.70)	7.77 (1.90)
BPVS standardised	84.43** (23.86)	70.56** (10.94)
TROG VMA (years)	6.51 (2.50)	6.29 (1.81)
TROG age-adjusted	106.33 (35.63)	94.35 (25.53)
Raven's CPM (raw score)	25.92** (7.42)	20.84** (7.01)
Ratio of ToM passers: failers	14: 23	15:17

Group comparisons \*  $p < .05$  \*\*  $p < .01$

The groups were constructed so as to be able to compare participants who failed FB tasks with those who passed. This meant that a disproportionately large number of the MLD group were FB 'failers' as compared to the percentage in the population at large, and a disproportionately large number of the ASD group were FB 'passers'. As a result, the MLD group has a significantly lower verbal IQ as measured by the BPVS. However, the groups were well matched on TROG VMA.

### **Procedure**

Each child was tested on the BPVS 2<sup>nd</sup> Edition (Dunn et al., 1999), the TROG (Bishop, 1989) and on Raven's CPM (Raven et al., 1998) in the course of the earlier studies. They were also given FB tasks as described in Study 1, and tested on the modified version of the WCST, as described in Study 2. Those who passed a majority of the FB tasks they were given were allocated to the 'FB passers' group. The remaining participants were allocated to the FB failers group.

### **Theory of Mind and Executive Function questionnaire**

This questionnaire was designed using 15 items from previous studies (e.g. Frith et al., 1994) that had been considered to require a ToM. These included telling white lies, taking things literally, recognising surprise or embarrassment and responding to indirect hints. Fifteen items were included that were designed to be

indicative of executive functioning (Booth et al., in prep) for example, having an ability to plan ahead, being able to do mental arithmetic and being able to follow verbally given lists of instructions. An additional five items were included which were social items that were thought not to necessitate a ToM, for example having appropriate table manners, and recognising happiness and sadness. These items went to form the 'general sociability' (G-S) scale. See Appendix II for a list of all items in the questionnaire.

The questionnaire was distributed, in most cases personally, to the form teachers of all the children who participated in studies 2, 3 and 4. The teachers were given the option of completing the questionnaire themselves, doing it as a face-to-face interview, or as a phone interview. The researcher's phone number and a stamped addressed envelope were included with the questionnaire. If teachers did not respond within 4 weeks, a reminder letter was sent, and if they did not respond within a further 4 weeks, a phone call was made to the school and in some cases further questionnaires were dispatched. In the case of the children who were recruited through volunteers, their parents and teachers were both given a copy of the questionnaire. If both were returned, the teacher's questionnaire was used in preference, but if the teacher did not return it then the parent's answers were taken (this happened in 4 cases, all of the children involved had ASD). Fifty-six questionnaires were completed individually by teachers (25 in the ASD group, 31 in the MLD group), often after consultation with the researcher, 8 (7 in the ASD group, 1 in the MLD group) were completed as face-to-face interviews, and one (in the ASD group) was completed as a phone interview. In total, 78% of the questionnaires distributed were completed (80% in the ASD group, 74% in the MLD group).

Scales: A scoring system similar to that used by Frith et al. (1994), based on that for the Vinelands Adaptive Behaviour Scales was utilised. If the respondent said the child could do a particular behaviour 'definitely', they were awarded a score of '2'. If they said the child could do it 'sometimes', 'rarely' or if they said they did not know, this scored '1'. If they said the child could do it 'not at all' then the child was awarded a score of '0' for that item. These scores were compiled to give 3 scales, two of which (the ToM and EF scales) ranged from 0-30, and one of which (G-S) ranged from 0-10. The alpha-values for the ToM scale was .81 in the ASD group and .78 in the MLD group, whilst for the EF scale they were .77 in the ASD group and .77 in the MLD group, all of which indicate very good levels of consistency. The alpha-values

for the G-S scale were .08 for the ASD group and .41 for the MLD group. The lower values for the G-S scale reflect the diverse nature of the items in this scale: there was no theoretical reason why these items should hang together other than by virtue of not being thought to require a ToM.

## **Results**

The diagnostic groups will be compared on their ratings on the questionnaire scales. Next, the ratings of FB passers and failers will be compared within each group, and then the relationship between the questionnaire scales and performance on the card sort task will be reported. Finally the role of ability and the relationship between the different questionnaire scales will be discussed.

### **Do the MLD and ASD groups differ on ratings of ToM, EF or general sociability?**

Table 9.2 gives the mean scores of the two groups on the 3 scales. A univariate ANOVA found a significant difference between the groups (with the MLD group outperforming the ASD group) on the ToM scale ( $F(1,65) = 4.02, p < .05$ ). There was no difference between the groups on the EF scale ( $F(1,65) = 0.93, p = .34$ ), or on the general sociability scale ( $F(1,65) = 0.10, p = .75$ ).

***Table 9.2 Mean score on the 3 questionnaire scales. Mean, (s.d.), range***

	ASD (N=37)	MLD (N=32)
ToM scale (max =30)	14.22*, (5.20), 4-25	16.75*, (5.02), 8-27
EF scale (max = 30)	12.65, (5.06), 5-25	13.78, (4.63), 7-23
G-S scale (max = 10)	6.76, (1.19), 4-9	6.66, (1.43), 4-9

Group comparisons, \* =  $p < .05$

### **Do those who pass or fail FB tasks differ on ratings of real life behaviour?**

Table 9.3 gives the mean scores on the scales for the ToM passers and failers as well as the verbal and non-verbal ability of each group. Figure 9.1 graphically illustrates the different scores for the subscales for MLD FB passers and failers, and ASD FB passers and failers.

One-way ANOVAs found differences between the passers and failers in both groups on all ability measures (ASD all  $F$ -values  $(1,35) > 11.4$ , all  $p$ -values  $< .01$ , MLD all  $F$ -values  $(1,30) > 8.5$  all  $p$ -values  $< .01$ ). The ASD group also differed on

age ( $F(1,35) = 5.90, p < .05$ ), and there was a trend towards an age difference in the MLD group ( $F(1,30) = 3.62, p = .07$ ).

**Table 9.3 Scores for FB passers and failers on age, verbal ability and non-verbal ability, and the ToM, EF and Sociability questionnaire scales (mean (s.d.) for ability and age, mean, (s.d.) range for scales)**

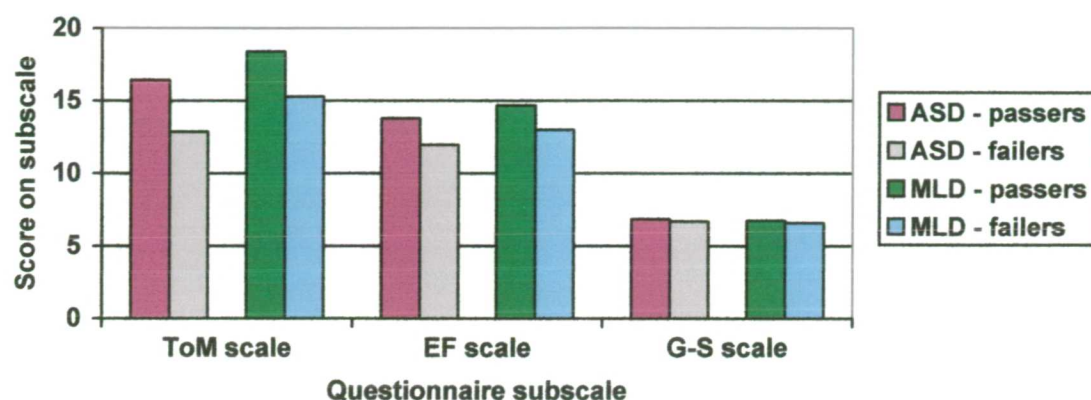
	ASD (N = 37)		MLD (N=32)	
	Passers (N = 14)	Failers (N = 23)	Passers (N=15)	Failers (N=17)
CA (years)	12.08* (1.45)	10.26 (2.55)	12.67 (.92)	11.73 (1.70)
BPVS VMA (years)	12.46** (3.22)	6.53 (1.53)	8.97** (1.00)	6.71 (1.88)
BPVS standardised	105.36** (24.15)	71.70 (11.77)	76.80** (5.91)	65.06 (11.51)
TROG VMA (years)	8.89** (2.39)	5.07 (1.03)	7.18** (1.77)	5.50 (1.48)
TROG age-adjusted	133.22** (28.19)	89.96 (29.46)	107.46** (23.19)	82.78 (22.11)
Raven's CPM raw score (max = 36)	30.57** (5.5)	23.09 (7.07)	26.00** (5.84)	16.29 (4.31)
ToM scale	16.43* (5.27) 4-25	12.87 (4.77) 4-20	18.40 (5.07) 11-27	15.29 (4.65) 8-24
EF scale	13.79 (6.20) 7-25	11.96 (4.23) 5-24	14.67 (5.54) 7-23	13.00 (3.66) 8-21
G-S scale	6.86 (1.35) 4-9	6.70 (1.11) 5-9	6.73 (1.39) 5-9	6.59 (1.50) 4-9

\*  $p < .05$ , \*\*  $p < .01$

ANOVAs showed significant differences between passers and failers in the ASD group on the ToM questionnaire scale only (ToM scale  $F(1,35) = 4.47, p < .05$ , EF scale  $F(1,35) = 1.14, p = .29$ , G-S scale  $F(1,35) = .16, p = .69$ ). There was a trend towards a difference in the MLD group on the ToM scale only (ToM  $F(1,30) = 3.36, p = .08$ , EF  $F(1,30) = 1.97, p = .17$ , G-S  $F(1,30) = .18, p = .68$ ).

In order to look at the relationship between passing or failing FB and performance on the various questionnaire subscales on a more individual level, the questionnaire ratings were split by the median score (15 for the ToM scale, 13 for the EF scale, 7 for the general sociability scale), and the number of participants scoring at or above this median score was calculated for each group (see table 9.4).

**Figure 9.1. Subscale scores for MLD passers and failers, and ASD passers and failers. The ToM and EF scales range from 0-30, whilst the general sociability scale ranges from 0-10**



**Table 9.4 Number of participants in each group who scored at or above the median for the questionnaire subscales**

		ASD at or above median (N passers = 14, N failers = 23)	$\chi^2$ (df), sig.	MLD at or above median (N passers = 15, N failers = 17)	$\chi^2$ (df), sig.
ToM scale	FB Failers	15/8	6.68 (1), p=.01	9/8	2.28 (1), p=.13
	FB Passers	3/11		4/11	
EF scale	FB Failers	12/11	0.30 (1) p =.58	9/8	0.54 (1), p=.35
	FB Passers	6/8		6/9	
G-S scale	FB Failers	11/12	0.52 (1), p = .47	8/9	0.00 (1), p=.98
	FB Passers	5/9		7/8	

The chi-square tests showed there was a strong association between performing at or above the median on the ToM scale and passing FB in the ASD group only. However, it was not the case that all the FB failers did badly on the scale, 8/23 (35%) received ratings above or at the median. In addition, 21 % (3/14) of the ASD FB passers received ratings below the median on the ToM scale.

### **Does performance on a test of executive function relate to ratings of real life behaviour?**

All participants were given a modified version of the WCST (as described in Study 2, Chapter 5). Table 9.5 gives the mean scores for the two diagnostic groups for percentage perseverative errors, percentage conceptual level sorting, no. of trials to complete first category, failure to maintain set and no. of categories achieved.

Table 9.6 gives the correlations of the individual variables of the card sort with the questionnaire subscales. A card sort aggregate was also created by taking the means for the 5 card sort variables of the entire sample, assigning a value of 2 to those who were above the 75<sup>th</sup> percentile, 0 to those who were below the 25<sup>th</sup>, and 1 to those between the 25<sup>th</sup> and 75<sup>th</sup>. These scores were then summed to give a score that could potentially range from 0-10. This aggregate therefore gave participants a score relative to other participants in the sample. Correlations with this variable are given in the last column of Table 9.6.

**Table 9.5. Mean scores for the card sort variables for the ASD and MLD groups; mean (s.d.)**

	ASD (N = 37)	MLD (N=32)
Percentage perserverative errors	28.01 (27.35)	24.44 (18.55)
Percentage conceptual level sorting	43.57 (30.71)	40.50 (26.09)
No. of trials to complete first category	9.71 (4.76)	8.96 (3.51)
Failure to maintain set (no. of times)	0.11 (0.31)	0.41 (0.61)
No. of categories achieved.	2.65 (1.46)	2.47 (1.44)
Card sort aggregate	5.40 (2.84)	4.72 (2.82)

Both the ToM and the EF questionnaire scales related to the Card Sort task – most significantly in the MLD group, but with a small to moderate effect size in the ASD group as well. This relationship was particularly with the percentage conceptual level processing and the number of categories achieved, both measures of general success on the task. It seemed likely that these associations could be due to the relationship of ability with both the questionnaire subscales and performance on the card sort tasks, and so Raven’s Matrices (as a non-verbal ability measure) was partialled out of the correlations. These values are given in brackets in table 9.6. The relationship between the ToM scales and conceptual level sorting remained significant in both diagnostic groups, whilst the relationship with the EF scale (which had been significant in the MLD group only) fell below significance in the MLD group and became much lower in both groups.



**Table 9.6 Correlations between questionnaire subscales and card sort variables in the ASD and MLD groups (values with Raven's partialled out). ASD in bold.**

		Percentage perseverative errors	Percentage conceptual level processing	Failure to maintain set	Number of trials to complete first category	Number of categories achieved	Card sort aggregate
ASD (N=37)	ToM scale	<b>-.24</b> <b>(-.19)</b>	<b>.39*</b> <b>(.39*)</b>	<b>-.07</b> <b>(-.11)</b>	<b>-.30</b> <b>(-.23)</b>	<b>.12</b> <b>(.15)</b>	<b>.33*</b> <b>(.33)</b>
	EF scale	<b>-.11</b> <b>(.03)</b>	<b>.26</b> <b>(.13)</b>	<b>-.08</b> <b>(-.10)</b>	<b>-.05</b> <b>(.05)</b>	<b>.05</b> <b>(-.13)</b>	<b>.21</b> <b>(.05)</b>
	G-S scale	<b>-.08</b> <b>(-.31)</b>	<b>.11</b> <b>(.20)</b>	<b>.22</b> <b>(.26)</b>	<b>.09</b> <b>(.15)</b>	<b>-.02</b> <b>(.13)</b>	<b>-.02</b> <b>(-.02)</b>
MLD (N=32)	ToM scale	<b>-.30</b> <b>(-.19)</b>	<b>.45*</b> <b>(.39*)</b>	<b>.01</b> <b>(-.11)</b>	<b>-.25</b> <b>(-.23)</b>	<b>.39*</b> <b>(.15)</b>	<b>.36*</b> <b>(.33)</b>
	EF scale	<b>-.31</b> <b>(.03)</b>	<b>.46**</b> <b>(.13)</b>	<b>-.09</b> <b>(-.10)</b>	<b>-.39</b> <b>(.05)</b>	<b>.36*</b> <b>(-.13)</b>	<b>.43*</b> <b>(.05)</b>
	G-S scale	<b>-.30</b> <b>(-.31)</b>	<b>.22</b> <b>(.20)</b>	<b>.05</b> <b>(.26)</b>	<b>-.21</b> <b>(.15)</b>	<b>.21</b> <b>(.13)</b>	<b>.18</b> <b>(-.02)</b>

\*p < .05, \*\*p < .01

### **The role of ability**

The ToM passers and failers in each group differed significantly on verbal and non-verbal ability, as did the MLD and ASD groups. This is a possible cause of differences on the questionnaire scales. However, ToM task performance is so closely linked with verbal ability in ASD that it does not make sense to control for this factor, since if the FB passers and failers were equivalent in verbal ability they would in all likelihood no longer be FB passers and failers.

In order to look at the relationship between ability and performance on the questionnaire scales without controlling for these inherent group differences, the relationship between ability and performance on the questionnaire scales was assessed separately within each group (ASD passers and failers, MLD passers and failers). This was done by a series of simple linear regression analyses in each group successively, using Raven's Matrices as an estimate of non-verbal ability. Non-verbal ability did not predict performance on the ToM scale in any of the groups (all  $R^2 < .11$ , all F values < 1.5, all p-values > .23). On the EF scale, non-verbal ability predicted performance in the ASD passers only ( $R^2 = .30$ ,  $F = 2.29$ ,  $p < .05$ ). All other  $R^2 < .03$ , all other F values < 0.56, all p-values > .46). On the G-S scale, non-verbal ability did not predict performance in any of the groups (all  $R^2 < .02$ , all F values < 0.18, all p-values < .68). Therefore, it appeared that the relationship between performance on the

questionnaire scales and non-verbal ability was restricted to one scale in one group. This means that adjusting for differences in ability between groups is probably unnecessary in the case of the ToM and G-S scales (since there is no evidence for a relationship between performance and ability in the separate groups). It also makes any sort of controlling or adjusting for ability in the case of the EF scale problematic since an assumption of such group comparisons is that the linear relationship between the potential covariate (in this case non-verbal ability) and the outcome variable (in this case the EF scale) is the same, and in this case this assumption is not met.

The difference in non-verbal and verbal ability between the MLD and the ASD groups could also be considered to be problematic, since again it is an inherent part of the groups' identity – the ASD group is of higher ability because it includes a high proportion of individuals who pass ToM tasks, which in individuals with ASD is strongly associated with having high verbal ability. In contrast, the ability range of the MLD group is restricted by the very fact of their having moderate learning difficulties. However, this difference is not really problematic for the interpretation of the results, since the group difference on the ToM questionnaire scale between the ASD and MLD group goes in the opposite direction to that which would be predicted by the ability measures (i.e. the MLD group outperform the ASD group, despite the latter's an ability advantage). Therefore, were we to control for ability, we would expect the difference to increase in magnitude, rather than decrease.

### **How do the questionnaire subscales relate to ability, age and each other?**

An aim of this study was to look at the relationship between real life ratings of ToM and EF, as opposed to the relationship between experimental ToM and EF tasks. Table 9.7 gives the correlations between the ability measures and performance on the 3 subscales of the questionnaire for the two groups.

With the G-S scale partialled out (as a measure of tendency of the teacher to rate a certain child particularly high or particularly low), the correlations between the ToM and EF scales remained significant in both groups (ASD  $r(37) = .48, p < .01$ ; MLD  $r(32) = .63, p = .001$ ).

The correlation between the ToM and EF scales also remain significant with CA, BPVS VMA, TROG VMA and Raven's CPM were partialled out in addition to the G-S scale (ASD  $r(37) = .52, p < .005$ , MLD  $r(32) = .57, p < .005$ ). It is possible that this relationship is an artefact of some participants having relatively good EF and

ToM, whilst others perform poorly on both (rather than being due to a linear relationship between EF and ToM). In order to look at this, correlations were calculated separately for the FB passers and failers in each group

**Table 9.7 Correlations between the questionnaire subscales and ability measures ASD group in bold).**

		EF scale	G-S scale	CA (yrs)	BPVS VMA	BPVS std	TROG VMA	TROG std	Ravens
ASD (N=37)	ToM scale	<b>.55**</b>	<b>.32</b>	<b>.03</b>	<b>.22</b>	<b>.24</b>	<b>.32</b>	<b>.32</b>	<b>.12</b>
	EF scale		<b>.40*</b>	<b>.13</b>	<b>.21</b>	<b>.24</b>	<b>.14</b>	<b>.09</b>	<b>.19</b>
	G-S scale			<b>.21</b>	<b>.10</b>	<b>.09</b>	<b>.15</b>	<b>.06</b>	<b>.03</b>
MLD (N=32)	ToM scale	<b>.68**</b>	<b>.33</b>	<b>.12</b>	<b>.45**</b>	<b>.51**</b>	<b>.34</b>	<b>.44</b>	<b>.23</b>
	EF scale		<b>.46*</b>	<b>.17</b>	<b>.30</b>	<b>.27</b>	<b>.16</b>	<b>.30</b>	<b>.24</b>
	G-S scale			<b>-.38</b>	<b>-.12</b>	<b>.17</b>	<b>-.01</b>	<b>.18</b>	<b>.05</b>

\* p<.05, \*\* p<.01

. In the ASD failers, but not the passers, the correlation between the EF and ToM scales was significant (failers  $r(23) = .63$ ;  $p < .001$ ; passers  $r(14) = .42$ ;  $p = .13$ ). In the MLD group, both passers' and failers' ToM and EF ratings were correlated (failers  $r(17) = .74$ ;  $p < .001$ ; passers  $r(15) = .63$ ;  $p < .05$ ). When the G-S scales, age and ability were all partialled out, these correlations were: ASD failers  $r(23) = .64$ ,  $p < .01$ , ASD passers  $r(14) = .18$ ,  $p = .65$ , MLD failers  $r(17) = .56$ ,  $p = .06$ , MLD passers  $r(15) = .60$ ,  $p = .06$ . Therefore the relationship appears to hold up in all groups except for the ASD passers, with large effect sizes in all three of the other groups (Cohen, 1969).

## **Discussion**

### **Do the ASD and MLD groups differ on real life measures of ToM, EF and general sociability?**

The children in this study were in some ways exceptional. The ASD group had an uncharacteristically high percentage of individuals who passed FB tasks, whilst the MLD group had a particularly high percentage of individuals who failed. Therefore, the fact that a significant difference was found between the ASD and MLD groups (with the MLD group outperforming the ASD group) on the ToM scale of the questionnaire is striking, particularly since the ASD group had an advantage in non-

verbal ability (and of verbal IQ) over the MLD group. This is in line with the previous experimental findings of this thesis (Study 2: Chapter 5), which found that even when only children who fail FB tasks were included in the study, children with MLD still performed better on a range of other ToM tasks than did children with ASD.

There were no differences on measures of executive functioning in everyday life between the ASD and MLD groups. This again is in line with the experimental findings reported in chapter 5, which found no major group differences in performance on EF tasks amongst those who failed FB tasks (although there were differences in the card sort once non-verbal ability was accounted for). This may indicate that both children with ASD and those with MLD have real life executive difficulties, or that neither do. Without normal controls it is impossible to tell, however it is clear that the data indicate that there was no specific ASD deficit in real life executive functioning as measured by the present questionnaire. This may seem to contradict the findings of Turner (1997). However, the present questionnaire did not assess repetitive behaviour and as such was quite different to that used by Turner.

There were also no differences on a general sociability measure, indicating that individuals with ASD are not necessarily pervasively impaired on all social activities, but appear to have particular difficulties with those that require a ToM. This is in line with other studies using questionnaire ratings (e.g. Frith et al., 1994), which have found that children with ASD were unimpaired on sociability items which do not require mentalising.

### **How do ratings on scales of ToM and EF relate to performance on experimental tasks?**

There was a relationship between passing and failing FB tasks and ratings on the ToM scale of the questionnaire in the ASD group only. FB passers performed significantly better than failers on this scale, and a higher proportion of them performed above the median score. There was a trend towards a difference on the ToM scales between FB passers and failers in the MLD group, but this did not reach significance and there was no difference in the proportion of passers and failers who performed above the median. As mentioned in the results section, the groups differed significantly on ability, and it is a matter of some debate whether this should be controlled for in some way. At least one previous study has controlled for a group

difference (between ToM passers and failers) in verbal ability and has found that group differences on questionnaire ratings are no longer evident (Fombonne et al., 1994). In the present thesis, it was decided for theoretical reasons that the group ability difference was so fundamental to the nature of the groups that it would be unreasonable to attempt to control for this difference. It is not possible to envisage a group of FB passers with ASD who are matched in ability to a group of FB failers – they would in all probability no longer be groups of passers and failers. By the same token, it is fundamental to the nature of the MLD group that they are of lower ability than the ASD group, since if they were of higher ability they would no longer be described as having learning difficulties.

General measures of success on the card sort task were related to ratings on the ToM and EF questionnaire scales in both the ASD and MLD groups. Once non-verbal ability was partialled out the relationship was only significant with the ToM scale. This may indicate that EF as measured experimentally is important for the expression of every day ToM – it may be, for example, that a lack of flexibility limits the ability of children to shift between perspectives and conversational topics. Turner (1997) found a relationship between questionnaire ratings of everyday EF and experimental performance, however her questionnaire scales were much more focused on repetitive behaviour, and it may be that the EF scale used in the present study was too general to relate to specific EF tasks. In a future, it would be helpful to measure more specific elements of everyday EF - for example, having subscales to measure working memory, planning, or flexibility, rather than the general scale used in this study.

There was a very robust relationship between the EF and ToM questionnaire scales, which held up even when age, ability and the ‘general sociability’ scale were partialled out. This is in contrast to the experimental findings of chapter 5, which found no relationship between performance on EF and ToM tasks in those who failed FB. One possibility for this inconsistency was that including both FB passers and failers may create the illusion of a linear relationship because some children perform well on both, whilst others performance badly on both. To test this, the relationship between the EF and ToM questionnaire ratings was assessed separately in FB passers and failers. There was a large effect size in all groups except for the ASD FB passers. Therefore it seems that the inclusion of both passers and failers is not necessary for this relationship to be evident, and indeed that the relationship is not restricted to

higher-functioning individuals (as it was suggested in chapter 5 might be the case with experimental measures), but might even be more evident in those who are lower functioning. This may be preliminary evidence for a relationship between the two in real life.

## **Conclusion**

The findings from this study replicate those of Frith et al. (1994) who found a significant difference on teacher ratings of real life ToM between individuals with ASD who passed and failed tests of FB understanding. In contrast to Frith et al. (1994), however, there was some tentative evidence in this study that real life ratings of ToM are related to FB task performance in children with MLD as well as in those with ASD. Frith et al. (1994) identified three subgroups of children with ASD: those who failed FB tasks and perform poorly on ratings of everyday mentalising ability; those who passed FB tasks but still show poor performance on ratings of everyday ToM; and those who passed FB and show superior performance on ratings of everyday ToM. In this study a fourth group was evident, those who failed FB tasks but showed relatively good performance on everyday ratings of ToM. This group is puzzling. Do they have some understanding of mind that is underestimated by experimental tasks – or have they worked out some strategies that work in every day life but which do not help on experimental tasks? There is much evidence that TD young children, for example, have mentalising abilities that are underestimated by ToM tasks (e.g. Bartsch & Wellman, 1995). However, children with ASD are quite different to TD young children, and their mentalising abilities are generally not thought to be underestimated by their frequent failure on FB tasks (e.g. Leslie, 1994). It is possible to envisage rule-based strategies that would not generalise to experimental situations with dolls, but which are useful in some real life situations. However, the items on the questionnaire were specifically designed to require flexible socialisation rather than fixed routines, which is what would be expected if children were using a rule-based strategy. There was no clue from this study as to how these different subgroups of children with ASD might be distinguished – ability did not relate to performance on the ToM scale, for example.

It is possible that the existence of FB failers who performed well on everyday mentalising measures is an artefact of the way that the data were collected. The passers and failers in the ASD group in this study came from almost entirely different

schools, and therefore the teachers may have had different baselines to which they were comparing them. In a school where most of the children have very low social functioning, a child with a slightly higher level of language and therefore some capacity to interact may be scored disproportionately highly, since the contrast to the remainder of the group is so stark. In contrast, in a school where most children have good language levels and some degree of social interaction, the comparisons are likely to be made with a different baseline level of what would be expected. This is in contrast to the MLD group, all of whom came from schools with both passers and failers. In the Frith et al. (1994) study, all the children came from schools with both FB passers and failers. Further studies, using other methods to assess everyday life ToM and EF are needed.

The overall picture from this study was one of the ASD group performing very badly in relation to the MLD group specifically on ratings of ToM (as opposed to EF or general sociability) in everyday life, despite an overall ability advantage, and regardless of their performance on experimental tasks. In addition, whether a child passed or failed tests of FB was related to their ratings on everyday mentalising ability in the ASD group. There was no real evidence for a relationship between performance on an EF task and ratings of everyday EF, although there did seem to be a relationship in both the ASD and MLD groups between experimental measures of EF and ratings of real life mentalising ability. The ToM and EF ratings were related in all the groups except for the ASD passers. There was evidence that a group of children with ASD may exist who fail FB tasks but are rated relatively highly for ToM in everyday life, although this may be an artefact of the teacher ratings and needs further investigation. This could be done through observational methods, or multi-report questionnaires rather than relying on a single teacher. It might also be useful to assess whether *environmental factors*, such as *intervention programmes* that schools may be using, have an impact on adaptation in the ASD group, and may account for the strong performance of some of those who fail FB tasks.

## **CHAPTER 10**

### **DISCUSSION**

This thesis has investigated a number of topics relating to ToM in individuals with ASD and MLD. This discussion chapter will briefly summarise the most important findings, and then will offer a theoretical perspective on these results and suggestions for future work.

#### **Summary of Results**

All five studies in this thesis compared children with ASD to those with MLD. Study 1 looked at the relationship between language and performance on FB tasks in a large sample of children. It found a highly significant difference between the proportion of children with ASD and MLD who passed FB tasks, with a higher proportion of children with ASD failing. It also found a strong relationship between language and performance on FB tasks, particularly in those with ASD and particularly with receptive grammar. Study 2 looked at the relationship between EF and ToM in a sample of children who failed FB tasks. It found no relationship between EF and ToM in either ASD or MLD, and no pervasive deficit in EF in the ASD group as compared to the MLD group, although the ASD group had a tendency towards making more perseverative errors.

Study 3 was a training study of ToM and EF, aiming to examine the direction of the relationship between the constructs. Children with both ASD and MLD were successfully trained on an alternative strategy for ToM, which generalised to a deception task in both groups, and to other tasks in the MLD group. ToM training did not improve performance on EF tasks. The EF training programme (which focused on set shifting, inhibitory control and flexibility) did not improve performance on executive tasks, however, those trained showed a significant improvement in ToM two months later.

Study 4 looked at children with ASD who passed FB tasks, asking whether their performance on a range of advanced ToM tasks provided clues as to how they might be achieving their relative success on FB tasks. The performance of these high-



functioning children with ASD was remarkably strong across a range of ToM tasks. However, they were still impaired (when compared to VMA matched controls) on reflecting on their own thoughts, and on attributing mental states in response to non-verbal stimuli. There was evidence that some individuals with ASD may have been using a post hoc strategy for attributing mental states, rather than computing them online, and there was very tentative evidence that some children with ASD may be using verbal strategies to solve FB tasks.

Finally, Study 5 looked at teacher ratings of everyday behaviour in children with ASD and MLD. Despite groups that contained an abnormally high level of FB passers (in the ASD group) and FB failers (in the MLD group), there was still a significant difference between the groups on ratings of behaviour requiring mental state attribution, with the ASD group showing less of this type of behaviour than the MLD group. This was despite a VMA advantage in the ASD group. There was no difference on ratings of behaviour requiring executive control, or on a general sociability measure. Those individuals with ASD who passed FB tasks performed significantly better on teacher ratings of behaviour requiring mental state attributions than those who failed. In the MLD group there was a trend towards a similar relationship. Again, there were no differences on the other questionnaire scales. Ratings of real life behaviour requiring mental state attribution and ratings of behaviour requiring executive control were strongly and robustly related in both groups.

## **Discussion of Results**

### **Theory of Mind in children with ASD**

One of the aims of this thesis was to look at the relative degree of impairment in ToM in children with ASD and MLD, and to look for factors other than genuine variation in the understanding of mind that might account for FB task success or failure.

This thesis found strong evidence that ToM deficits in children with ASD were both more common and more extensive than in children with MLD. When children who failed standard FB tasks were compared on a range of ToM tasks, those with ASD were more severely impaired than those with MLD. Those who failed FB tasks did not appear to be failing due to executive demands of the tasks, since

performance on EF tasks did not relate to performance on ToM tasks. In addition, FB task performance was related to ratings of everyday social insight.

From my reading of the literature (see chapter 1), it appears that all the current theories of ToM postulate the existence of some sort of predisposition to, or an innate basis for ToM, and many suggest that those with ASD have a deficit in this innate basis. Given this, the existence of a subgroup of individuals with ASD who pass ToM tasks is intriguing. It might be tempting to write this group off as simply having developed task-specific tactics that enable them to pass. However, previous findings of their superior everyday life social insight and their success on a range of other experimental tasks (see chapter 8), suggests that their success is due to something more than luck or a task-specific strategy, and therefore merits further investigation.

What then, is the nature of the understanding of mind in children with ASD? There are two main possibilities. They may be delayed in their acquisition of a ToM, or their development of the understanding of mind may be qualitatively different to that seen in other groups. This is harder to test than one might think, since we know very little about normal socio-cognitive development of ToM over the age of 5, and the discrepancy in ability levels means that tasks that seem appropriate for high-functioning children with ASD are not appropriate for TD 5- or 6-year-olds. In addition, Fisher et al. (submitted) found that, in a sample of TD 8-year-olds, variation in performance on advanced ToM tasks (such as the Strange Stories or the ToM cartoons) appeared to reflect differences in language ability rather than genuine differences in the understanding of mind, indicating that elucidating normal social development may not as straightforward as giving advanced ToM tasks to a range of ages of normally developing children. There may well be quite a narrow window in normal development when a particular advanced ToM test actually assesses variation in the understanding of mind, as opposed to other, more general processes. Having a metric of the developmental progression of performance on tests of understanding of mind would enable us to identify exactly where children with ASD are relative to their typically developing peers, and whether their socio-cognitive development is taking a delayed or deviant pathway.

Given the present paucity of information about the normal development of social cognition over the age of 5 years, is there any other way in which we could test for delayed, rather than deviant, ToM ability? Variation in performance across tasks may be informative, and the results of study 4 suggest that the ASD group did show a

different profile of strengths to that seen in the MLD group. The ASD group did not show deficits in everything when compared even to the VMA matched MLD group – there were no problems on the Strange Stories (Happé, 1994), for example, or in defining mental states. However, they did display deficits on attributing mental states to non-verbal stimuli, and reporting back their own thoughts. Study 5 suggested that those ‘FB passers’ with ASD were most comparable to ‘FB failers’ with MLD on ratings of real life behaviour requiring mental state attribution, again suggesting that the development of understanding of mind in ASD is not simply delayed. Perhaps the most convincing argument against the delay hypothesis is that high-functioning individuals with ASD do not appear to ever quite catch up, and most remain clearly ‘autistic’ throughout life – as were even the most able and oldest children with ASD in these studies.

The alternative is that any understanding of mind gained by children with ASD is quite different in nature to that shown by TD children and children with MLD. It is possible that individuals with ASD have an innate deficit underlying their difficulties in representing mental states, which will inevitably mean that their social development is different to TD children. In the case of high-functioning children with ASD, the notion of ‘child as scientist’ may well be appropriate, with children having very little or no innate basis for or predisposition towards understanding other minds, and instead having to painstakingly construct a theory of representational mental states. For children with ASD, unlike TD children, the predictions from such an account in terms of individual differences seems to hold; not all children with ASD come up with a ToM at the same time (and in fact many never do), and there is no reason to assume that those who do have all formed the same theory. Subgroups of children with ASD may exist, with differing profiles of social strengths and weaknesses, according to the particular understanding of mind that they have constructed. This of course makes research into their difficulties much more complicated, and it may be that the only way forward is to use much larger samples of children with ASD who pass FB tasks, in order to identify subgroups who may have specific patterns of performance on ToM tasks. Alternatively it might be possible to select subgroups who demonstrate particular difficulties in real life behaviour and to group them together, looking for characteristic deficits on experimental tasks.

## **Language and ToM**

Much of the evidence of this thesis points to one thing; language appears to be crucially important in the development of an understanding of mind in children with ASD. Language ability (and particularly grammar) predicted who would pass and fail FB tasks, and related strongly to performance on more advanced ToM tasks in those who did pass FB.

Given that language appears to be a very likely candidate for an alternative route to ToM in able individuals with ASD (Happé, 1995; Tager-Flusberg, 2002), there are two main theories as to the route that this might take. Language may provides access to social relationships and conversations, which provide the child with experience of and subsequently an understanding of mental states (Peterson & Siegal, 1995; Dunn & Brophy, 2002), or specific aspects of language such as complement syntax may provide a child with the structure within which they can represent other perspectives (De Villiers, 2000). The findings reported in this thesis suggest that grammar is, indeed, especially important to the development of ToM in ASD (see studies 1 and 4). The measure in this study did not include complement syntax, and this may indicate that the relationship is in fact more general than that suggested by De Villiers, or that the grammar measures used may in fact have operated as a proxy measure of complement syntax ability. This point could be clarified by including more extensive measures of grammar in the future, as well as looking specifically at which aspects of grammar children with ASD who fail FB tasks find hard. However, it appears unlikely that an association with complement syntax can account for the entire relationship seen between grammar and ToM in this thesis. This is because the relationship is just as evident in the 'FB passers' group, all of whom passed 1<sup>st</sup> order FB, and therefore presumably could use complement syntax. The importance of grammar suggests that improved access to conversations and social interactions alone cannot account for the relationship with FB since vocabulary should surely be equally important in that case (as, in fact, it appears to be in those with MLD), and provides some indication that the structure of language itself might be crucial. It is possible that the embedded structure even of non-mental states terms may be helpful to individuals with ASD in developing the ability to shift between perspectives – many of the later items of the TROG are embedded, for example 'the book the pencil is on is red', requires the child to hold in mind 'the book' whilst

listening to the rest of the sentence, and then to shift back to the subject in order to work out what the last clause ('is red') refers to.

In the case of children with ASD, who may have an impaired innate basis for the understanding of mind, language could provide them with a concrete example of other perspectives that they can begin to comprehend without necessarily having to represent mental states in an abstract way. This is because language makes mental states explicit, and can therefore emphasize the fact that different people have different perspectives, possibly enabling the person with ASD to form a concrete representation of what a person says (which could potentially be conceptualised almost like a speech bubble), rather than having to form an abstract representation of what they think. They therefore may be able to start constructing their own explicit theory of ToM, from their experience and learnt rules. This may have implications for the trajectory of the development of understanding of mind in ASD. In TD young children, it appears that they may first display implicit social understanding (for example, looking towards the place where the protagonist will search in a FB task) before they demonstrate an explicit understanding of mind (for example predicting correctly where the protagonist will look (e.g. Clements, Rustin, & McCallum, 2000)). It is possible that children with ASD show a very different learning curve, where mental states have to be made explicit before they can be understood at all, and where no implicit understanding precedes the explicit demonstration of understanding. This hypothesis could be tested by developing a range of tests where children with ASD's responses were monitored in a more implicit way. Study 4 of this thesis started to look at this, and found that some children with ASD failed an 'implicit' 1<sup>st</sup> order FB task despite passing 2<sup>nd</sup> order tasks when they were asked explicitly about the protagonists' belief. Other potential experimental methods might include eye tracking paradigms, as suggested by Klin, Jones, Schultz et al. (2002). In addition, if children with ASD have particular problems in the implicit understanding of other minds, they should continue to have difficulties in certain areas no matter how able they are. One such area could be reflecting on and reporting back their own mental states. There has been very little research on the ability to 'read own mind' in individuals with ASD. Frith and Happé (1999) suggested that self consciousness and the ability to report on an internal stream of consciousness may be impaired in even high-functioning individuals with ASD, and that the degree of impairment may be related to performance on FB tasks. Study 4 of this thesis (with FB passers only) included a task

that asked children with ASD to reflect on their thoughts, and on which they performed remarkably badly when compared to the VMA matched MLD group, despite good performance on FB tasks and other ToM tasks. In the future it would be useful to extend this finding by comparing FB passers and failers on their ability to report back their thoughts, as well as developing more tasks that ask high-functioning individuals with ASD to reflect on their stream of consciousness.

It is possible that the need for mental states to be made explicit in order for children with ASD to learn about them explains the relative success of the ToM training programme reported in this thesis (Study 3). The ToM programme made the concept of differing perspectives and mistaken belief explicit, and provided the children with a linguistic structure within which to solve FB tasks. The real life improvement seen in over 50% of the trained children in Study 5 (in contrast to none of the control children) suggests, tentatively, that training programmes may in some cases have the potential to lead to genuine changes in the understanding of mind. Whether this has any lasting impact on the adaptive behaviour of children with ASD remains to be seen. Longer term intervention programmes are needed, perhaps carried out by teachers or parents over a period of months, rather than weeks, and reinforced by talking about the strategy in different contexts. In order to improve the possibility of generalisation of learning, the training should be repeated across different contexts (Koegel, Koegel, & Parks, 1995; Schreibman & Koegel, 1996). The fact that those children with ASD who pass FB tasks do show superior mentalising abilities in real life when rated by their teachers (Study 5, Fombonne et al., 1994; Frith et al., 1994) suggests that targeting these abilities explicitly could potentially be beneficial for real life adaptation.

Findings reported in this thesis suggest that language is not just important developmentally, but also remains crucially important for performance on ToM tasks online. Study 4 found that in some advanced ToM tasks high-functioning children with ASD showed an high degree of competence. Their understanding of mental states appeared to be accessed more readily through certain, language based routes – providing a narrative, for example, seemed to make ToM tasks easier for children with ASD, and tasks which could be answered by learnt rules about mental states also appeared to be relatively easy. Other studies have found a relationship between narrative ability and ToM in children with ASD, whilst finding no relationship in children with MLD (Capps, Losh, & Thurber, 2000), suggesting again that there may

be a special relationship between the two in ASD. Areas that remained impaired were reporting back their own mental states, attributing mental states in response to a non-verbal stimulus, and possibly the attributing of mental states to pictures of people's eyes. All these are areas in which it is hard to see how a ToM that remains essentially language bound could help. There was some very tentative evidence that verbal interference tasks may selectively impair performance on FB tasks in children with ASD, however, this needs to be repeated with a bigger sample and a more difficult ToM task. In order to examine this effect, it would be useful to develop tasks that look directly at the effect of narrative of ToM task performance on ASD. In particular, if identical versions of a videoed task were constructed, one with a voice over narrative, one without, then the impact of an explicit verbal narrative on performance could be directly assessed. The prediction would be that it would enhance performance on ToM in those with ASD, whilst having minimal effect on those with MLD, or possibly even confusing them and thus impairing performance. Control (non-ToM) stories would be necessary in case a narrative simply makes all tasks easier. This might be the case if, for example, explicit narratives helps individuals with ASD to extract global meaning.

The present study of those individuals with ASD who passed FB tasks suffered from a number of limitations apart from its small sample size, in particular the unmatched nature of the MLD and ASD groups. In the future it would be useful to conduct studies of this type with two control groups, one of which contains children with MLD (who would be likely to be of lower VMA), and one of TD individuals who could be matched on verbal age. However, as discussed in chapter 8, this would present a challenge in terms of age-appropriate task design.

It was not possible in this thesis to compare those with Asperger's Syndrome with those with autism, since there appeared to be no substantive difference according to teacher reports between those with the different diagnoses. However, in the future it would be useful to compare those individuals (matched on present day VMA) who had a language delay in their early years with those who did not, and to examine the implications of early language delay for later ToM development in ASD. One possibility is that early language gives a head start in the development of an alternative route to ToM, and therefore that participants with Asperger's Syndrome (as currently defined), will demonstrate significantly better ToM skills when compared to VMA matched individuals with high-functioning autism.

### **Theory of Mind in children with MLD**

A proportion of children with MLD are almost invariably found to fail FB tasks (e.g. Baron-Cohen et al., 1985; Charman & Shmueli-Goetz, 1998; Zelazo et al., 1996), and it is not clear whether this indicates a genuine difficulty in understanding other minds, or whether it is instead due to other, more general difficulties such as EF deficits. Some authors have suggested that for a child with MLD to pass or fail a FB task tells us relatively little about their underlying social competence (Frith et al., 1994; Happé, 1995), whilst others argue that failing these tasks is evidence of genuine ToM deficits in these groups, and therefore the lack of specificity of a ToM deficit to autism (e.g. Yirmiya et al., 1998).

Unlike previous work (e.g. Happé, 1995), studies in this thesis found a relationship between language and ToM in children with MLD (studies 1 and 4). The nature of this relationship was different to that seen in ASD. Whilst in ASD grammar appeared to be a significant and independent predictor of ToM, in MLD the relationship with vocabulary was as important as that with grammar, and grammar generally did not predict performance over and above vocabulary. In addition, in study 4, the non-ToM control tasks were as closely linked to language as the ToM tasks, indicating that task specific factors, rather than the ToM demands, may account for the relationship.

Somewhat surprisingly, there was some indication in this thesis that a subset of children with MLD may indeed have genuine difficulties in representing mental states. They responded in a very similar way to the children with ASD to the ToM and EF training programmes, and there was a trend towards those who failed FB tasks performing worse on ratings of everyday behaviour requiring mental state attributions than those who passed. In addition, they appeared to find a ‘memory of own FB’ task easier than predicting another’s FB, even though the non-ToM demands of the tasks were virtually identical; which indicates that these tests may have been taxing their mentalising abilities rather than other, more general, skills. However, this thesis also provided evidence that children with MLD who fail FB tasks are qualitatively different to children with ASD who fail. For a start, task-specific factors appeared to be more important in MLD, with a clear hierarchy of difficulty emerging between different FB tasks (study 1). In addition, even those participants with MLD who failed FB performed better on a range of ToM tasks than did children with ASD



(study 2), and were most comparable to those children with ASD who *passed* FB on ratings of every day behaviour requiring mental state attributions (study 5).

Given that it seems possible that a small subset of individuals with MLD have difficulties in FB which are indicative of some wider problems in ToM, it may be interesting to speculate on why this might be so. Children with MLD do not have the qualitative social and communicative impairments characteristic of autism, and we would assume that they have an intact innate basis for ToM, whatever this might be, in the same way as TD children. However, they are different to TD children in one crucial way; they have a low IQ and therefore a lower capacity for learning. All the major theories of ToM, as well as agreeing that there must be some innate basis for ToM, allow for the role of experience in the development of a true representational understanding of mind. By definition, children with MLD have difficulties in gathering and using information, and it is possible that for some of them this is also true in the social cognitive domain. Their ability to learn from their experiences may therefore be seriously delayed in some cases, and it may take much more exposure to conversations, social interactions, mistaken beliefs, deception and so forth in order for them to learn enough to pass a FB test. It is also plausible that, due to their slow development, they do not become ready to learn about representational mental states until they are much older than TD children. This delay could have serious implications for their development, since the information available in their environment when they *are* ready to learn may be quite different to that which they would have got as young children. Stories we tell to small children are full of simple scenarios of mistaken belief and deception (e.g. Red Riding Hood thinking the wolf is her grandmother, or Cinderella's sisters thinking she is a princess at the ball), and young children are typically cared for in small groups with lots of adult contact, giving them a particularly rich opportunity for individual interactions with those who are more socially sophisticated than themselves, but yet who are keen to instruct them on social norms and situations. It is possible that, having not been ready to learn about representational mental states when they were the 'right' age, some children with MLD never get such a rich environment for learning again. By the time they are ready to learn they are already at school in larger groups, being taught academic skills rather than the socio-cognitive understanding typically learnt in early childhood.

Evidence for lack of experience being a potential source of ToM problems in MLD comes from training studies which provide intensive experience of FB

situations, and which have generally been successful in children with MLD. Children with MLD and other non-autistic developmental disorders appear to be quite effectively trained on ToM, showing better generalisation than those with ASD (Study 3, Swettenham, 1996), maintenance of improvement over time (Study 3, Ashcroft et al., 1999) and, in one notable case, improvement on ratings of social skills in everyday life (Steerneman et al., 1996). In order to elucidate the true nature of ToM in children with MLD who fail FB tasks, more measures of everyday life behaviour would be useful, not just relying on teacher ratings but maybe also observing behaviour in dyadic interactions or family situations. This is another area where a clearer understanding of the normal developmental trajectory of the understanding of mind would be useful, in order to compare if and how those with MLD differ from the norm, if indeed they do. I would predict that individuals with MLD who fail FB tasks should be competent in a wide range of social cognitive skills, in a similar way to TD 3-year-olds, despite failing FB tasks. However, they may still display specific difficulties in some areas – for example in attributing mistaken beliefs that are not made explicit.

### **Executive Function and Theory of Mind**

The findings from this thesis concerning EF and ToM are mixed. One of the more puzzling findings was the success of the EF training programme in improving performance on ToM tasks, whilst not improving performance on EF tasks. As argued in Study 3, it is unlikely that non-specific training effects accounted for the improvement on ToM tasks, so it seems that something about the EF training must have led to the improvement; this is reinforced by the finding that those who improved on ToM scored more highly on ratings of success during the EF training than those who did not improve. If the study were to be repeated, it would be useful to have a wider range of executive tasks, some of which should be aimed at the same developmental level as FB tasks, and including other EF tasks that are known to be related to FB in normal development (e.g. the Windows task; (Russell et al., 1991). Study 2 found no association between EF and ToM in those who failed FB tasks (with either ASD or MLD), whilst Study 5 found a strong association between real life ratings and measures of ToM and EF, making the situation potentially even more puzzling. This is another area (as well as ToM, discussed above) where it would be useful to have a metric of the normal developmental trajectory of performance on the

tasks, in order to assess the nature of the difficulties in children with ASD. It would also have been useful to have included a range of EF tasks in the FB passers study in this thesis.

The literature to date suggests that there is a strong relationship between performance on FB tasks and EF tests measuring flexibility and inhibitory control such as card sort tasks (see chapter 4 for a summary). It appears possible that the relationship between ToM and EF is quite specifically linked to the transition to a representational understanding of mind. First order FB tasks require children to shift their perspective from reality to a mental state, whether their own past belief or another's present belief. It seems plausible that this shifting requires a degree of cognitive flexibility and working memory – the ability to 'hold in mind' one state of affairs (e.g. reality or own belief) whilst shifting to another (e.g. the protagonist's false belief). Findings from typically developing children do indeed show that this type of task is related to performance on FB tasks between the ages of 3 and 5 (see chapter 4). This is not to say that the relationship reflects only task specific factors, since this shifting between perspectives is also necessary in order to consider differing mental states in everyday life. When designing and carrying out the EF and ToM training programmes, it became clear how much both the target tasks (FB and Card Sort) were about shifting – the EF training programme was about shifting between actions, whilst the ToM programme was about shifting between mental states. The results of study 3 indicate that training children in cognitive shifting may have implications for their later ability to pass ToM tasks. The fact that the improvement seen in the EF trained groups in study 3 was more evident at follow-up suggests that the relationship between set shifting and ToM is not an immediate, task-specific, one. The ability to shift between actions (which most of the children were able to learn to do in the EF training programme) may enable children to begin to shift between perspectives and thus, through further experience (which may well be linguistically mediated), be able to develop some representational understanding of differing mental perspectives. Those who improved on ToM had better grammar at the start than those who did not, suggesting that grammatical ability and the training may have interacted in some way. The shifting that is necessary for an improved understanding of mental states may be quite different to the shifting that is measured in most EF tasks, which could explain why the children did not improve on the EF tasks.

It is also possible that the ability to hold in mind is particularly important in children with ASD, since (as argued above) they may be constructing an alternative route to ToM, which would probably require more cognitive resources than the intuitive (possibly modular) route taken in typical development. I would predict that a certain degree of cognitive flexibility would be necessary but not sufficient for the development of understanding of mind in ASD, that without some cognitive flexibility able children with ASD would be unable to make use of the explicit social experience and language ability that they do have. It is, however, also possible that without some cognitive flexibility individuals with ASD are not able to gain a good level of grammar, since, as discussed above, more complicated grammatical structures often contain embedded phrases and therefore require holding in mind and shifting between perspectives. One might predict that whilst children with ASD exist who have relatively good EF but poor ToM (study 2; Baron-Cohen & Robertson, 1995), the opposite pattern should not exist (poor EF, good ToM). This is only the case for those with ASD, since other children are assumed to have an intact basis for ToM, and to therefore be more able to make use of implicit cues, and therefore an explicit, cognitive strategy is less crucial in the development of their understanding of mind.

This idea is highly speculative, and in order to test it more studies are needed which compare those who pass and fail FB tasks on EF measures. The results of study 2 suggest that it is possible that there is no relationship between the understanding of mind and EF in those who fail FB tasks. It is likely that, whatever the pattern of development, it could be quite different in developmentally delayed groups, certainly some recent evidence indicates that there may not be a relationship in those with ADHD (Charman et al., 2001; Perner et al., 2002). This supports my earlier argument that cognitive flexibility may be less crucial in groups who have an innate ToM basis and who therefore have a more typical route to ToM. If the relationship between EF and ToM is due to the need to shift between perspectives in both domains, then we would not expect to see a relationship in those who fail FB tasks (as in fact study 2 found), since a certain level of EF should be necessary, but not sufficient, for a ToM. In those individuals with ASD who pass FB tasks, I would predict that a relationship between cognitive flexibility, working memory and ToM should be particularly evident, since they are hypothesised to be using an alternative strategy for the understanding of mind, which may require the individual to both 'hold something in mind' and to shift between different perspectives. In support of this, study 4 (with FB

passers) found that working memory was related to ToM in individuals with ASD only. In TD children, and those with non-autistic developmental disorders, I would predict that the relationship between EF and ToM would be evident over the period of the transition to FB understanding, when the individual may be learning to shift between conflicting mental perspectives for the first time. However, once they have achieved this milestone the shifting between perspectives become more automatic, and cognitive flexibility and working memory should no longer be crucial in the way that it remains for children with ASD. An important caveat is that the ability to shift between different frameworks, and good language ability, are not themselves enough to ensure that a child with ASD will work out some representational understanding of mental states; their experience of social interactions, which may be mediated by motivation, personality and interest, will also be essential.

## **General conclusions**

This thesis has compared the ToM abilities of those with MLD and those with ASD, and the relationship of ToM to EF and language. The evidence in those with MLD is tentative, but there was some evidence that passing or failing a FB task is not simply artefactual in their case, as had been suggested in the past. Language, particularly grammar, appears to be strongly developmentally implicated in ToM, particularly in ASD, and suggestions have been made for how and why this might be the case. In contrast, EF does not appear to be related to ToM in those individuals with MLD or ASD who failed FB tasks. Training children in ToM was reasonably effective and showed some generalisation in both ASD and MLD. Training children in cognitive flexibility and set shifting did not improve their EF, but seemed to improve their ToM, especially at follow-up. This improvement over time suggests that the EF training may work indirectly, possibly enabling children to learn more from social experiences and therefore improve their understanding of mind. It was suggested that language, cognitive flexibility and working memory may all be crucial in enabling children with ASD to construct and use a working understanding of mind. How exactly this process might work, and how language and executive function might interact is as yet unclear. Passing FB tasks appears to be indicative of a genuinely better understanding of mind in some children with ASD. The question of how they achieve this, and whether other children with ASD can be helped to achieve

this, is one that is still open, but which this thesis has made some effort to begin to explore.

## REFERENCES

- Abell, F., Happe, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. Cognitive Development, 15(1), 1-16.
- APA. (1994). Diagnostic and Statistical Manual of Mental Disorders (4th ed.). Washington: American Psychiatric Association.
- Appleton, M., & Reddy, V. (1996). Teaching three year-olds to pass false belief tests: A conversational approach. Social Development, 5(3), 275-291.
- Ashcroft, A., Jervis, N., & Roberts, C. (1999). A theory of mind (TOM) and people with learning disabilities: The effects of a training package. Journal of Applied Research in Intellectual Disabilities, 12(1), 58-68.
- Astington, J., & Jenkins, J. M. (1999). A Longitudinal Study of the Relation Between Language and Theory-of-Mind development. Developmental Psychology, 35(5), 1311-1320.
- Avis, J., & Harris, P. L. (1991). Belief-desire reasoning among Baka children: Evidence for a universal conception of mind. Child Development, 62(3), 460-467.
- Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder: Evidence from a British twin study. Psychological Medicine, 25, 63-77.
- Baird, G., Charman, T., Baron-Cohen, S., Cox, A., Swettenham, J., Wheelwright, S., & Drew, A. (2000). A screening instrument for autism at 18 months of age: A 6-year follow-up study. Journal of the American Academy of Child & Adolescent Psychiatry, 39(6), 694-702.

Baldwin, D., & Saylor, M. (2002). Language Promotes Structural Alignment in the Acquisition of a Theory of Mind. Paper presented at the Why Language Matters for Theory of Mind International Conference, Toronto, Canada.

Baron-Cohen, S. (1989). The autistic child's theory of mind: a case of specific developmental delay,. Journal of Child Psychology and Psychiatry, 30, 285-298.

Baron-Cohen, S. (1992). Out of sight or out of mind: another look at deception in autism. Journal of Child Psychology & Psychiatry & Allied Disciplines, 33, 1141-1155.

Baron-Cohen, S. (1995). Mindblindness: An essay on autism and theory of mind. Cambridge: MA: MIT Press.

Baron-Cohen, S. (2000). Theory of mind and autism: a fifteen year review. In D. J. Cohen (Ed.), Understanding Other Minds (2nd ed.). New York: Oxford University Press.

Baron-Cohen, S. (2001). Theory of mind and autism: A review. In L. M. E. Glidden (Ed.), International review of research in mental retardation: Autism (Vol. vol. 23, pp. pp. 169-184). San Diego: US: Academic Press.

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. British Journal of Psychiatry, 161, 839-843.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger Syndrome. Journal of Child Psychology and Psychiatry and Allied Disciplines, 38(7), 813-822.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? Cognition, 21(1), 37-46.



Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. British Journal of Developmental Psychology, 4(2), 113-125.

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children with asperger syndrome or high-functioning autism. Journal of Autism & Developmental Disorders, 29((5)), 407-418.

Baron-Cohen, S., & Robertson, M. M. (1995). Children with either autism, Gilles de la Tourette syndrome or both: Mapping cognition to specific syndromes. Neurocase, 1(2), 101-104.

Baron-Cohen, S., Spitz, A., & Cross, P. (1993). Can children with autism recognise surprise? Cognition and Emotion, 7, 507-516.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. Journal of Child Psychology & Psychiatry & Allied Disciplines, 42(2), 241-251.

Baron-Cohen, S., Wheelwright, S., Scahill, V., Spong, A., & Lawson, J. (in press). "Are intuitive physics and intuitive psychology independent? A test with children with Asperger Syndrome,". Journal of Developmental and Learning Disorders.

Baron-Cohen, S., Wheelwright, S., Stone, V., & Rutherford, M. (1999). A mathematician, a physicist and a computer scientist with Asperger syndrome: Performance on folk psychology and folk physics tests. Neurocase, 5(6), 475-483.

Bartsch, K., & Wellman, H. M. (1995). Children talk about the mind. Oxford: Oxford University Press.

Bettelheim, B. (1956). Childhood schizophrenia as a reaction to extreme situations. Journal of Orthopsychiatry, 26, 507-518.

Bishop, D. V. (1989). Test for reception of grammar. Cambridge: MRC Applied Psychology Unit.

Blackshaw, A. J., Kinderman, P., Hare, D. J., & Hatton, C. (2001). Theory of mind, causal attribution and paranoia in Asperger syndrome. Autism, 5(2), 147-163.

Bloom, P. (2000). How children learn the meaning of words. Cambridge, MA, US: The MIT Press.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. Cognition, 77(1), 25-31.

Booth, R., Happé, F., Hughes, C., & Charlton, R. (in prep). Executive function and theory of mind in everyday life.

Bosacki, S., & Astington, J. W. (1999). Theory of mind in preadolescence: Relations between social understanding and social competence. Social Development, 8(2), 237-255.

Bosacki, S. L. (2000). Theory of mind and self-concept in preadolescents: Links with gender and language. Journal of Educational Psychology, 92(4), 709-717.

Bowler, D. (1997). Reaction times to mental state and non-mental state questions in false belief tasks by high-functioning individuals with autism. European Child & Adolescent Psychiatry, 6(3), 160-165.

Bowler, D. M. (1992). "Theory of mind" in Asperger's syndrome. Journal of Child Psychology & Psychiatry & Allied Disciplines, 33(5), 877-893.

Bowler, D. M., & Briskman, J. A. (2000). Photographic cues do not always facilitate performance on false belief tasks in children with autism. Journal of Autism & Developmental Disorders, 30(4), 305-316.

Bowler, D. M., & Strom, E. (1998). Elicitation of first-order "theory of mind" in children with autism. Autism, 2(1), 33-44.

Capps, L., Kehres, J., & Sigman, M. (1998). Conversational abilities among children with autism and children with developmental delay. Autism, 2, 325-344.

Capps, L., Losh, M., & Thurber, C. (2000). "The frog ate the bug and made his mouth sad": Narrative competence in children with autism. Journal of Abnormal Child Psychology, 28(2), 193-204.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. Child Development, 72(4), 1032-1053.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the Relation between Executive Function and Theory of Mind? Contributions of Inhibitory Control and Working Memory. Infant and Child Development, 11, 73-92.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social Competence, Joint Attention and Communicative Competence from 9 to 15 months of Age. Monographs of the Society for Research in Child Development, 63(4).

Chakrabarti, S., & Fombonne, E. (2001). Pervasive developmental disorders in preschool children. Journal of the American Medical Association, 285(24), 3093-3099.

Charman, T., Baron-Cohen, S., Baird, G., Cox, A., Wheelwright, S., Swettenham, J., & Drew, A. (2001). Commentary: The Modified Checklist for Autism in Toddlers. Journal of Autism & Developmental Disorders, 31(2), 145-148.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2001). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. Cognitive Development, 15(4), 481-498.

Charman, T., Carroll, F., & Sturge, C. (2001). Theory of mind, executive function and social competence in boys with ADHD. Emotional and Behavioural Difficulties, 6(1), 31-49.

Charman, T., & Lynggaard, H. (1998). Does a photographic cue facilitate false belief performance in subjects with autism? Journal of Autism & Developmental Disorders, 28(1), 33-42.

Charman, T., Ruffman, T., & Clements, W. (2002). Is there a gender difference in false belief development? Social Development, 11(1), 1-10.

Charman, T., & Shmueli-Goetz, Y. (1998). The relationship between theory of mind, language and narrative discourse: an experimental study. Cahiers de Psychologie Cognitive, 17(2), 245-271.

Clarke, A. M., & Clarke, A. D. B. (1976). Formerly isolated children. In A. D. B. Clarke (Ed.) (pp. 27-34). London: Open Books.

Clements, W., Rustin, C., & McCallum, S. (2000). Promoting the transition from implicit to explicit understanding: a training study of false belief. Developmental Science, 3(1), 81-92.

Cohen, J. (1969). Statistical Power Analysis for the Behavioral Sciences. London: Academic Press, Inc.

Cohen, N., Vallance, D. D., Barwick, M., Im, N., Menna, R., Horodezky, N. B., & Isaacson, L. (2000). The Interface between ADHD and Language Impairment: An Examination of Language, Achievement, and Cognitive Processing. Journal of Child Psychology and Psychiatry, 41(3), 353-362.

Cole, K., & Mitchell, P. (2000). Siblings in the development of executive control and a theory of mind. British Journal of Developmental Psychology, 18(2), 279-295.

Coltheart, M. (1999). Modularity and Cognition. Trends in Cognitive Sciences, 3(3), 115-120.

Corcoran, R. (2000). Theory of mind in other clinical conditions: is a selective 'theory of mind' deficit exclusive to autism. In D. J. Cohen (Ed.), Understanding other Minds: Perspectives from Developmental Cognitive Neuroscience (pp. 391-421). Oxford, UK: Oxford University Press.

Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language and family background: Individual differences and inter-relations. Child Development, 70(4), 853-865.

Dahlgren, S. O., & Trillingsgaard, A. (1996). Theory of Mind in Non-Retarded Children with Autism and Asperger's Syndrome: A Research Note. Journal of Child Psychology and Psychiatry, 37(6), 759-763.

Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. Australian Journal of Psychology, 47(1), 25-31.

De Villiers, J. (2000). Language and Theory of Mind: what are the developmental relationships? In D. J. Cohen (Ed.), Understanding other Minds: Perspectives from Developmental Cognitive Neuroscience (2nd ed., pp. 83-123). Oxford: Oxford University Press.

De Villiers, J. (2002). Can language acquisition give children a point of view? Paper presented at the Why Language Matters for Theory of Mind, Toronto, Canada.

De Villiers, J., & Pyers, J. E. (2002). Complements to cognition: a longitudinal study of the relationship between complex syntax and false belief understanding. Cognitive Development, 17, 1037-1060.

Drewe, E. A. (1975). Go-no go learning after frontal lobe lesions in humans. Cortex, 11(1), 8-16.

Dunn, J., & Brophy, M. (2002). Conversation, Relationships and the Understanding of Mind. Paper presented at the Why Language matters for Theory of Mind International Conference, Toronto, Canada.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. Child Development, 62(6), 1352-1366.

Dunn, L. M., & Dunn, L. M. (1981). Peabody Picture Vocabulary Test-Revised. Circle Pines, MN: American Guidance Services.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1999). The British Picture Vocabulary Scale. Windsor: NFER-Nelson.

Fisher, N., Dunn, J., & Hughes, C. (submitted). Advanced tests of theory of mind: Are appearances deceptive? Findings from a study of 8-year-olds.

Flavell, J. H. (1999). Cognitive Development: Children's Knowledge about the Mind. Annual Review of Psychology, 50, 21-45.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1-Level 2 distinction. Developmental Psychology, 17(1), 99-103.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1995). Young children's knowledge about thinking. Monographs of the Society for Research in Child Development, 60(1).

Flavell, J. H., Green, F. L., & Flavell, E. R. (2000). Development of Children's Awareness of Their Own Thoughts. Journal of Cognition and Development, 1, 97-112.

Flavell, J. H., Green, F. L., Flavell, E. R., & Grossman, J. B. (1997). The development of children's knowledge about inner speech. Child Development, 61, 915-928.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & Frith, C. D. (1995). Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition, 57(2), 109-128.

Fombonne, E., Siddons, F., Achard, S., Frith, U., & et al. (1994). Adaptive behaviour and theory of mind in autism. European Child & Adolescent Psychiatry, 3(3), 176-186.

Frith, U. (1989). Autism: Explaining the enigma. Oxford, UK: Blackwell Scientific.

Frith, U., & Happé, F. (1999). Theory of mind and self-consciousness: What is it like to be autistic? Mind & Language, 14(1), 1-22.

Frith, U., Happé, F. G. E., & Siddons, F. (1994). Autism and theory of mind in everyday life. Social Development, 3, 108-124.

Frith, U., Morton, J., & Leslie, A. M. (1991). The cognitive basis of a biological disorder: Autism. Trends in Neurosciences, 14(10), 433-438.

Frye, D. (1999). Development of intention: The relation of executive function to theory of mind, Zelazo, Philip David (Ed); Astington, Janet Wilde (Ed); et al. (1999). Developing theories of intention: Social understanding and self-control. (pp. 119-132). Mahwah, NJ, US; Mahwah, NJ, US: Lawrence Erlbaum Associates, Inc., Publishers; Lawrence Erlbaum Associates, Inc., Publishers.

Frye, D., Zelazo, P. D., & Burack, J. A. (1998). Cognitive complexity and control: I. Theory of mind in typical and atypical development. Current Directions in Psychological Science, 7(4), 116-121.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. Cognitive Development, 10(4), 483-527.

Gale, E., de Villiers, P., de Villiers, J., & Pyers, J. (1996). Language and theory of mind in oral deaf children. In A. Zukowski (Ed.), Proceedings of the 20th annual Boston University Conference on Language Development (Vol. 1). Somerville, MA: Cascadilla Press.

Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of the mind' in verbal and nonverbal tasks. Neuropsychologia, 38(1) Jan 2000, 11-21.

Garner, C., Callias, M., & Turk, J. (1999). Executive function and theory of mind performance of boys with fragile-X syndrome. Journal of Intellectual Disability Research, 43(6), 466-474.

German, T. P., & Leslie, A. M. (2001). Children's inferences from 'knowing' to 'pretending' and 'believing.' British Journal of Developmental Psychology, 19, 59-83.

Gillberg, C., & Coleman, M. (2001). The Biology of Autistic Spectrum Disorders. Cambridge: Cambridge University Press.

Gopnik, A., Capps, L., & Meltzoff, A. (2000). Early theories of mind: what the theory theory can tell us about autism. In D. J. Cohen (Ed.), Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience (2nd ed.). Oxford: Oxford University Press.



Gopnik, A., & Wellman, H. M. (1994). The 'theory' theory. In L. A. Hirschfield, Gelman, S.A. (Ed.), Mapping the Mind: Domain Specificity in Cognition and Culture (pp. 257-293). Cambridge, U.K.: Cambridge University Press.

Gordon, A. C., & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. Journal of Experimental Child Psychology, 68(1), 70-83.

Gordon, R. M. (1986). Folk Psychology as simulation. Mind and Language, 1, 158-171.

Grant, C. M., Grayson, A., & Boucher, J. (2001). Using tests of false belief with children with autism: how valid and reliable are they? Autism, 5(2), 135-145.

Griffith, E. M., Pennington, B. F., Wehner, E. A., & Rogers, S. J. (1999). Executive functions in young children with autism. Child Development, 70(4), 817-832.

Grodzinsky, G. M., & Barkley, R. A. (1999). Predictive power of frontal lobe tests in the diagnosis of attention deficit hyperactivity disorder. Clinical Neuropsychologist, 13(1), 12-21.

Hadwin, J., Baron-Cohen, S., Howlin, P., & Hill, K. (1996). Can we teach children with autism to understand emotions, belief, or pretence? Development & Psychopathology, 8(2), 345-365.

Hadwin, J., Baron-Cohen, S., Howlin, P., & Hill, K. (1997). Does teaching theory of mind have an effect on the ability to develop conversation in children with autism? Journal of Autism & Developmental Disorders, 27(5), 519-537.

Hale, C. M., & Tager-Flusberg, H. (2002). The Influence of Language on Theory of Mind: A Training Study. Developmental Science (in press).

Happé, F. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. Cognition, 48(2), 101-119.

Happé, F. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. Journal of Autism and Development Disorders, 24(2), 129-154.

Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. Child Development, 66(3), 843-855.

Happé, F., Brownell, H., & Winner, E. (1999). Acquired "theory of mind" impairments following stroke. Cognition, 70(3), 211-240.

Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., Dolan, R., Frackowiak, R., & Frith, C. (1996). 'Theory of mind' in the brain. Evidence from a PET scan study of Asperger syndrome. Neuroreport, 8(1), 197-201.

Harris, E. L., Schuerholz, L. J., Singer, H. S., Reader, M. J., & et al. (1995). Executive function in children with Tourette syndrome and/or attention deficit hyperactivity disorder. Journal of the International Neuropsychological Society, 1(6), 511-516.

Harris, P. L. (1992). From simulation to folk psychology: the case for development. Mind and Language, 7, 120-144.

Harris, P. L., & Martin, L. (unpublished). From Little Red Riding Hood to Othello: Empathizing with a naive protagonist.

Hautus, M. J., & Lee, A. J. (1998). The dispersions of estimates of sensitivity obtained from four psychophysical procedures: implications for experimental design. Perception & Psychophysics, 60(4), 638-649.

Heaton, R. K. (1981). Wisconsin card sorting test manual. Odessa, FL: Psychological Assessment Resource Inc.

Hobson, R. P. (2002). The Cradle of Thought: Exploring the Origins of Thinking. London: Macmillan.

Hogrefe, J. G., Wimmer, H., & Perner, J. (1986). Ignorance versus False Belief: a developmental lag in attribution of epistemic states. Child Development, 57(3), 567-582.

Hughes, C. (1998a). Executive function in preschoolers: Links with theory of mind and verbal ability. British Journal of Developmental Psychology, 16, 233-253.

Hughes, C. (1998b). Finding your marbles: Does preschoolers' strategic behaviour predict later understanding of mind? Developmental Psychology, 34(6), 1326-1339.

Hughes, C. (2001). Executive dysfunction in autism: Its nature and implications for the everyday problems experienced by individuals with autism, Burack, Jacob A. (Ed); Charman, Tony (Ed); et al. (2001). The development of autism: Perspectives from theory and research. (pp. 255-275). Mahwah, NJ, US: Lawrence Erlbaum Associates, Inc., Publishers.

Hughes, C., & Cutting, A. L. (1999). Nature, Nurture and Individual Differences in Early Understanding of Mind. Psychological Science, 10(5), 429-432.

Hughes, C., Dunn, J., & White, A. (1998). Trick or treat?: uneven understanding of mind and emotion and executive dysfunction in "hard-to-manage" preschoolers. Journal of Child Psychology & Psychiatry & Allied Disciplines, 39(7), 981-994.

Hughes, C., & Russell, J. (1993). Autistic children's difficulty with mental disengagement from an object: Its implications for theories of autism. Developmental Psychology, 29, 498-510.

Hughes, C., Russell, J., & Robbins, T. W. (1994). Evidence for central executive dysfunction in autism. Neuropsychologia, 32, 477-492.

Hughes, C., Soares-Boucaud, I., Hochmann, J., & Frith, U. (1997). Social behaviour in pervasive developmental disorders: Effects of informant, group and "theory of mind." European Child & Adolescent Psychiatry, 6(4), 191-198.

Jenkins, J. M., & Astington, J. W. (1996). Cognitive Factors and Family Structure Associated With Theory of Mind Development in Young Children. Developmental Psychology, 32(1), 70-78.

Kanner, L. (1943). Autistic Disturbances of Affective Contact. The Nervous Child, 2, 217-250.

Kazak, S., Collis, G.M. & Lewis, V. (1997). Can young people with autism refer to knowledge states? Evidence from their understanding of 'know' and 'guess'. Journal of Child Psychology and Psychiatry, 38, 1001-1009.

Kerns, K. A., Eso, K., & Thomson, J. (1999). Investigation of a direct intervention for improving attention in young children with ADHD. Developmental Neuropsychology, 16(2), 273-295.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Defining and quantifying the social phenotype in autism. American Journal of Psychiatry, 159(6), 909-916.

Klin, A., Schultz, R., & Cohen, D. J. (2000). Theory of mind in action: developmental perspectives on social neuroscience. In D. J. Cohen (Ed.), Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience (2 ed.). Oxford: Oxford University Press.

Knoll, M., & Charman, T. (2000). Teaching false belief and visual perspective taking skills in young children: Can a theory of mind be trained? Child Study Journal, 30(4), 273-304.

Koegel, R. L., Koegel, L. K., & Parks, D. R. (1995). "Teach the individual" model of generalization: Autonomy through self-management. In L. K. Koegel (Ed.), Teaching children with autism: Strategies for initiating positive interactions and improving learning opportunities (pp. 67-77). Baltimore, MD: Paul H.

Kreppner, J. M., O'Connor, T. G., Dunn, J., & Andersen-Wood, L. (1999). The pretend and social role play of children exposed to early severe deprivation. British Journal of Developmental Psychology, 17(3), 319-332.

Leekam, S. R., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? Cognition, 40(3), 203-218.

Leslie, A., & Roth, D. (1994). What autism teaches us about metarepresentation. In S. E. Baron-Cohen, Tager-Flusberg Helen (Ed); et al. (Ed.), Understanding other minds: Perspectives from autism (Vol. . pp. 83-111). New York, NY, US: Oxford University Press.

Leslie, A. M. (1987). Pretense and representation: the origins of 'theory of mind'. Psychological Review, 94, 412-426.

Leslie, A. M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In S. A. Gelman (Ed.), Mapping the Mind: Domain Specificity in Cognition and Culture (pp. 119-148). Cambridge, UK: Cambridge University Press.

Leslie, A. M., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. British Journal of Developmental Psychology, 6(4), 315-324.

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. Developmental Science, 1(2), 247-253.

Lohmann, H., & Tomasello, M. (2002). The Role of Language in the Development of False-Belief Understanding: A Training Study. Paper presented at the Why Language Matters for Theory of Mind International Conference, Toronto, Canada.

Lord, C., & Pickles, A. (1996). Language Level and Nonverbal Social-Communicative Behaviors in Autistic and Language-Delayed Children. Journal of the American Academy of Child and Adolescent Psychiatry, 35(11), 1542-1550.

Loveland, K. A., & Landry, S. H. (1986). Joint Attention and Language in Autism and Developmental Language Delay. Journal of Autism and Developmental Disorders, 16(3), 335-349.

Lundy, J. E. B. (2002). Age and language skills of deaf children in relation to theory of mind development. Journal of Deaf Studies & Deaf Education, 7(1), 41-56.

Luria, A., Pribram, K. H., & Homskaya, E. D. (1964). An experimental analysis of the behavioural disturbance produced by a left frontal arachnoidol endothelioma (meningioma). Neuropsychologia, 2, 257-280.

MacGregor, E., Whiten, A., & Blackburn, P. (1998). Teaching theory of mind by highlighting intention and illustrating thoughts: A comparison of their effectiveness with 3-year-olds and autistic individuals. British Journal of Developmental Psychology, 16, 281-300.

McEvoy, R. E., Rogers, S. J., & Pennington, B. F. (1993). Executive function and social communication deficits in young autistic children. Journal of Child Psychology & Psychiatry & Allied Disciplines, 34(4), 563-578.

McGregor, E., Whiten, A., & Blackburn, P. (1998a). Teaching theory of mind by highlighting intention and illustrating thoughts: A comparison of their effectiveness with 3-year olds and autistic individuals. British Journal of Developmental Psychology, 16(3), 367-387.

McGregor, E., Whiten, A., & Blackburn, P. (1998b). Transfer of the picture-in-the-head analogy to natural contexts to aid false belief understanding in autism. Autism, 2(4), 367-387.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. Developmental Psychology, 31(5) Sep 1995, US,.

Miller, C. A. (2001). False belief understanding in children with specific language impairment. Journal of Communication Disorders, 34(1-2), 73-86.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding Analysis of Covariance. Journal of Abnormal Psychology, 110(1), 40-48.

Minshew, N., Goldstein, G., & Siegal, D. J. (1995). Speech and Language in High-Functioning Autistic Individuals. Neuropsychology, 9(2), 255-261.

Minter, M., Hobson, R. P., & Bishop, M. (1998). Congenital visual impairment and 'theory of mind.' British Journal of Developmental Psychology, 16(2), 183-196.

Mitchell, P., & Lacohee, H. (1991). Children's Early Understanding of false belief. Cognition, 39, 107-128.

Morgan, A. B., & Lilienfeld, S. O. (2000). A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. Clinical Psychology Review, 20(1), 113-136.

Mottron, L., & Burack, J. A. (2001). Enhanced perceptual functioning in the development of autism, Burack, Jacob A. (Ed); Charman, Tony (Ed); et al. (2001). The development of autism: Perspectives from theory and research. (pp. 131-148). Mahwah, NJ, US: Lawrence Erlbaum Associates, Inc., Publishers.

Mundy, P., Sigman, M., & Kasari, C. (1990). A Longitudinal Study of Joint Attention and Language Development in Autistic Children. Journal of Autism and Developmental Disorders, 20(1), 115-128.

Mundy, P., & Stella, J. (2000). Joint attention, social orienting, and nonverbal communication in autism, Wetherby, Amy M. (Ed); Prizant, Barry M. (Ed). (2000). Autism spectrum disorders: A transactional developmental perspective. Communication and language intervention series; vol. 9. (pp. 55-77). Baltimore, MD, US: Paul H. Brookes Publishing Co.

Ozonoff, S. (1997). Components of executive function in autism and other disorders. In J. E. Russell (Ed.), Autism as an executive disorder (pp. 179-211). New York: Oxford University Press;.

Ozonoff, S., & Jensen, J. (1999). Specific executive function profiles in three neurodevelopmental disorders. Journal of Autism & Developmental Disorders, 29(2), 171-177.

Ozonoff, S., & McEvoy, R. E. (1994). A longitudinal study of executive function and theory of mind development in autism. Development & Psychopathology, 6(3), 415-431.

Ozonoff, S., & Miller, J. N. (1995). Teaching theory of mind: A new approach to social skills training for individuals with autism. Journal of Autism & Developmental Disorders, 25(4), 415-433.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. Journal of Child Psychology & Psychiatry & Allied Disciplines, 32(7), 1081-1105.

Parsons, S., & Mitchell, P. (1999). What children with autism understand about thoughts and thought bubbles. Autism, 3(1), 17-38.



Pennington, B. F., & Ozonoff, S. (1996). Executive Functions and Developmental Psychopathology. Journal of Child Psychology and Psychiatry, 37(1), 51-87.

Perner, J., Kain, W., & Barchfield, P. (2002). Executive Control and Higher-Order Theory of Mind in Children at Risk of ADHD. Infant and Child Development, 11(2), 141-158.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. Trends in Cognitive Sciences, 3(9), 337-344.

Perner, J., & Lang, B. (2000). Theory of mind and executive function: Is there a developmental relationship? In D. J. Cohen (Ed.), Understanding other minds: Perspectives from developmental cognitive neuroscience (2nd ed., pp. 150-181). Oxford: Oxford University Press.

Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. Child Development, 73(3), 752-767.

Perner, J., Leekam, S., & Wimmer, H. (1987). Three-year-olds difficulty with false belief: The case for a conceptual deficit. British Journal of Developmental Psychology, 5, 123-137.

Perner, J., Stummer, S., & Lang, B. (1999). Executive functions and theory of mind: Cognitive complexity or functional dependence? In P. D. E. A. Zelazo, Janet Wilde (Ed); et al. (Ed.), Developing theories of intention: Social understanding and self-control. (pp. 133-152). Mahwah, NJ, US; Mahwah, NJ, US: Lawrence Erlbaum Associates, Inc., Publishers.

Perner, J., & Zauner, P. (2002). What does 'that' have to do with points of view? The case of 'want' in German? Paper presented at the Why Language Matters for Theory of Mind International Conference, Toronto.

Peterson, C. C., Peterson, J. L., & Webb, J. (2000). Factors influencing the development of a theory of mind in blind children. British Journal of Developmental Psychology, 18(3), 431-447.

Peterson, C. C., & Siegal, M. (1995). Deafness, Conversation and Theory of Mind. Journal of Child Psychology and Psychiatry, 36(3), 459-474.

Plaisted, K. C. (2001). Reduced generalization in autism: An alternative to weak central coherence. In J. A. Burack, Charman, Tony et al. (Ed.), The development of autism: Perspectives from theory and research (pp. 149-169).

Powell, J. E., Edwards, A., Edwards, M., Pandit, B. S., Sungum-Paliwal, S. R., & Whitehouse, W. (2000). Changes in the incidence of childhood autism and other autistic spectrum disorders in preschool children from two areas of the West Midlands, UK. Developmental Medicine & Child Neurology, 42(9), 624-628.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? Behaviour & Brain Sciences, 4, 515-526.

Rall, J., & Harris, P. L. (2000). In Cinderella's slippers? Story comprehension from the protagonist's point of view. Developmental Psychology, 36(2), 202-208.

Raven, J., Raven, J. C., & Court, J. H. (1998). Coloured Progressive Matrices. Oxford: Oxford Psychologists Press.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. Perceptual and Motor Skills, 8, 271-276.

Renfrew, C. E. (1991). The Bus Story: A test of narrative speech. Oxford, UK: Winslow Press.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. Developmental Psychology, 33(1), 12-21.

Roth, D., & Leslie, A. M. (1998). Solving belief problems: Toward a task analysis. Cognition, 66(1), 1-31.

Rumsey, J. M., & Hamburger, S. D. (1988). Neuropsychological findings in high-functioning men with infantile autism, residual state. Journal of Clinical and Experimental Neuropsychology, 10(201-221).

Russell, J. (1996). Agency: Its role in mental development. Hove, UK: Erlbaum (UK) Taylor and Francis.

Russell, J. (Ed.). (1997a). Autism as an executive disorder: New York, NY, US: Oxford University Press. (1997). viii, 313pp.

Russell, J. (1997b). How executive disorders can bring about an inadequate 'theory of mind'. In J. Russell (Ed.), Autism as an executive disorder (pp. 256-304). New York, USA: Oxford University Press.

Russell, J. (2002). Cognitive theories of autism. In A. M. Owen (Ed.), Cognitive Deficits in Brain Disorders. London: Martin Dunitz.

Russell, J., Jarrold, C., & Potel, D. (1994). What makes strategic deception difficult for children - the deception or the strategy? British Journal of Developmental Psychology, 12, 301-314.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The "windows task" as a measure of strategic deception in preschoolers and autistic subjects. British Journal of Developmental Psychology, 9, 331-349.

Rutter, M., Silberg, J., O'Connor, T. G., & Simonoff, E. (1999). Genetics and Child Psychiatry: II Empirical Research Findings. Journal of Child Psychology & Psychiatry & Allied Disciplines, 40, 19-55.

Scholl, B. J., & Leslie, A. M. (2001). Minds, modules, and meta-analysis. Commentary on "Meta-analysis of theory-of-mind development: The truth about false belief." Child Development, 72(3), 696-701.

Schreibman, L., & Koegel, R. L. (1996). Fostering self-management: Parent-delivered pivotal response training for children with autistic disorder. In P. S. Jensen (Ed.), Psychosocial treatments for child and adolescent disorders: Empirically based strategies for clinical practice (pp. 525-552). Washington, DC: American Psychological Association.

Semrud-Clikeman, M., Nielsen, K. H., Clinton, A., Sylvester, L., Parle, N., & Connor, R. T. (1999). An intervention approach for children with teacher- and parent-identified attentional difficulties. Journal of Learning Disabilities, 32(6), 581-590.

Sergeant, J. A., Geurts, H., & Oosterlaan, J. (2002). How specific is a deficit of executive functioning for attention-deficit/hyperactivity disorder? Behavioural Brain Research, 130(1-2), 447-471.

Shah, A., & Frith, U. (1983). An islet of ability in autistic children: a research note. Journal of Child Psychology & Psychiatry & Allied Disciplines, 24, 613-620.

Shah, A., & Frith, U. (1993). Why do autistic individuals show superior performance on the Block Design task? Journal of Child Psychology & Psychiatry & Allied Disciplines, 34, 1351-1364.

Shields, J., Varley, R., Broks, P., & Simpson, A. (1996). Social Cognition in Developmental Language Disorders and High-Level Autism. Developmental Medicine and Child Neurology, 38, 487-495.

Shimmon, K., & Lewis, C. (2001). Can Children with Autism be Trained on the Tower of London task? Paper presented at the Poster presented at BPS Developmental and Educational Sections Conference, Worcester, UK.

Siegal, M., & Beattie, K. (1991). Where to look first for children's understanding of false beliefs. Cognition, 38, 1-12.

Sigman, M., & Ruskin, E. (1999). Continuity and Change in the Social Competance of Children with Autism, Down Syndrome and Developmental Delays. Monographs of the Society for Research in Child Development, 64(1).

Slaughter, V. (1998). Children's understanding of pictorial and mental representations. Child Development, 69(2), 321-332.

Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. Child Development, 67(6), 2967-2988.

Sparrevohn, R., & Howie, P. (1995). Theory of Mind in Children with Autistic Disorder: Evidence of Developmental Progression and the Role of Verbal Ability. Journal of Child Psychology and Psychiatry, 36(2).

Steerneman, P., Jackson, S., Pelzer, H., & Muris, P. (1996). Children with social handicaps: An intervention programme using a theory of mind approach. Clinical Child Psychology & Psychiatry, 1(2), 251-563.

Sullivan, K., & Winner, E. (1991). When 3-year-olds understand ignorance, false belief and representational change. British Journal of Developmental Psychology, 9(1), 159-171.

Swettenham, J. (1996). Can children be taught to understand false belief using computers? Journal of Child Psychology & Psychiatry & Allied Disciplines, 37(2), 157-165.

Swettenham, J. (2000). Teaching theory of mind to individuals with autism. In D. J. Cohen (Ed.), Understanding other Minds: Perspectives from Developmental Cognitive Neuroscience. Oxford: Oxford University Press.

Swettenham, J. G., Baron-Cohen, S., Gomez, J.-C., & Walsh, S. (1996). What's inside someone's Head? Conceiving of the Mind as a Camera Helps Children with Autism Acquire an Alternative to a Theory of Mind. Cognitive Neuropsychiatry, 1(1), 73-88.

Tager-Flusberg, H. (2000). Language and understanding minds: Connections in autism. In D. Cohen (Ed.), Understanding other Minds: Perspectives from Developmental Cognitive Neuroscience (2nd ed., pp. 124-149). Oxford: Oxford University Press.

Tager-Flusberg, H. (2001). A reexamination of the theory of mind hypothesis of autism. In J. A. E. C. Burack, Tony (Ed); et al. (Ed.), The development of autism: Perspectives from theory and research. (pp. 173-193). Mahwah, NJ, US: Lawrence Erlbaum Associates, Inc., Publishers.

Tager-Flusberg, H. (2002). How Language Facilitates the Acquisition of False Belief in Children with Autism. Paper presented at the Why Language Matters for Theory of Mind International Conference, Toronto, Canada.

Tager-Flusberg, H., & Sullivan, K. (1994). Predicting and Explaining Behavior: A Comparison of Autistic, Mentally Retarded and Normal Children. Journal of Child Psychology and Psychiatry, 35(6), 1059-1075.

Tager-Flusberg, H., & Sullivan, K. (1995). Attributing mental states to story characters: a comparison of narratives produced by autistic and mentally retarded individuals. Applied Psycholinguistics, 16, 241-256.

Tager-Flusberg, H., Sullivan, K., & Boshart, J. (1997). Executive functions and performance on false belief tasks. Developmental Neuropsychology, 13(4), 487-493.

Taylor, M., & Hort, B. (1990). Can children be trained in making the distinction between appearance and reality? Cognitive Development, 5(1), 89-99.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. Child Development, 57(6), 1454-1463.

Turner, M. (1997). Towards an executive dysfunction account of repetitive behaviour in autism, Russell, James (Ed). (1997). Autism as an executive disorder. (pp. 57-100). New York, NY, US; New York, NY, US: Oxford University Press; Oxford University Press.

Varley, R., Siegal, M., & Want, S. C. (2001). Severe impairment in grammar does not preclude theory of mind. Neurocase, 7(6), 489-493.

Watson, A. C., Nixon, C., Wilson, A., & Capage, L. (1999). Social Interaction Skills and Theory of Mind in Young Children. Developmental Psychology, 35(2), 386-391.

Wechsler, D. (1974). Wechsler Intelligence Scale for Children - Revised. New York: The Psychological Corporation.

Wellman, H. M., Baron-Cohen, S., Caswell, R., Gomez, J. C., Swettenham, J., Toye, E., & Lagattuta, K. (2001). Using Thought-Bubbles Helps Children with Autism Acquire an Alternative to a Theory of Mind. Paper presented at the Meetings of the Cognitive Development Society, Norfolk VA.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. Child Development, 72(3), 655-684.

WHO. (1993). The ICD-10 classification of mental and behavioural disorders. Diagnostic criteria for research. Geneva: World Health Organisation.

Wimmer, H. (1989). Common-sense Mentalismus und Emotion: einige entwicklungs-psychologische Implikationen. In E. Roth (Ed.), Denken und Fühlen (pp. 56-66). Berlin: Springer Verlag.

Wimmer, H., Hogrefe, G. J., & Perner, J. (1988). Children's understanding of informational access as source of knowledge. Child Development, 59(2), 386-396.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining functions of wrong beliefs in young children's understanding of deception. Cognition, 13, 103-128.

Wing, L. (1981). Sex ratios in early childhood autism and related conditions. Psychiatry Research, 5, 129-137.

Wing, L., & Gould, J. (1979). Severe Impairments of social interaction and associated abnormalities in children: epidemiology and classification. Journal of Autism & Developmental Disorders, 9, 11-29.

Woolfe, T., Want, S. C., & Siegal, M. (2002). Signposts to Development: Theory of Mind in Deaf Children. Child Development, 73(3), 768-778.

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. Psychological Bulletin, 124(3), 283-307.

Zahn-Waxler, C., Robinson, J. L., & Emde, R. N. (1992). The Development of Empathy in Twins. Developmental Psychology, 28(6), 1038-1047.

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. Cognition, 35(1), 41-68.

Zaitchik, D. (1991). Is only seeing really believing? Sources of the true belief in the false belief task. Cognitive Development, 6(1), 91-103.



Zelazo, P. D., Burack, J. A., Benedetto, E., & Frye, D. (1996). Theory of Mind and rule use in individuals with Down's Syndrome: A test of the uniqueness and specificity claims. Journal of Child Psychology & Psychiatry & Allied Disciplines, 37(4), 479-484.

Zelazo, P. D., Burack, J. A., Boseovski, J. J., Jacques, S., & Frye, D. (2001). A cognitive complexity and control framework for the study of autism. In J. A. E. C. Burack, Tony (Ed); et al. (Ed.), The development of autism: Perspectives from theory and research. (pp. 195-217). Mahwah, NJ, US: Lawrence Erlbaum Associates, Inc., Publishers.

Zelazo, P. D., Burack, J. A., & Frye, D. (2002). The Relation Between Theory of Mind and Rule Use: Evidence from Persons with Autism-Spectrum Disorders. Infant and Child Development, 11(2), 171-195.

Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. Review of General Psychology, 1(2), 198-226.

Zelazo, P. D., & Jacques, S. (2002). Why Does Language Matter for Theory of Mind? Because it Facilitates Self-Reflection and Cognitive Control. Paper presented at the 'Why Language Matters for Theory of Mind' International Conference, Toronto, Canada.

## **APPENDIX I SCRIPTS FOR FALSE BELIEF TASKS**

### **Deceptive box: Smarties Task (self and other)**

*First, can you tell me the name of a friend of yours?*

*I'm going to show you something I've brought.*

*What do you think is in here?*

If child responds "Other" say:

*"What do you normally find in here?"*

*"Okay, let's have a look... Open tube and show child the contents (pencils).*

*....There are pencils in here."*

*Right, let's put them back*

*Now if your friend (inset friend's name) came in here, what he will think is in here?*

Prompt if necessary: *What will (friend) think is in here?*

Control question: *What's really in the tube?*

Self question: *Before I opened the tube to show you, what did you think was in the tube?*

### **Unexpected Transfer: Sally-David Story**

*This is David and this is Sally. David has a bag, and Sally has a box.*

*David has an apple. David puts his apple away in his bag, to keep it safe while he plays outside. Then David goes out to play.*

*But look, while David is outside, Sally takes the apple out of David's bag, and she puts it in her box, here. Then Sally goes outside to play.*

*David comes in from playing - he's hungry and he'd like to have a bite of his apple.*

FB question: *Where will David look for his apple?*

Justification: *Why?*

Reality check control: *Where is the apple really?*

Memory check control: *Where was the apple first of all?*

## **APPENDIX II QUESTIONNAIRE ITEMS**

**(all items were made gender appropriate for the particular child).**

### **Theory of mind scale**

Does she draw people's attention to things that might interest them? For example, pointing out a type of car to someone who is interested in cars.

Does she deliberately try to pretend things, for example that she's ill, to get your sympathy or to get out of doing things?

Does she recognise when other people are surprised or embarrassed?

Does she sometimes misunderstand what is said to her because she takes it too literally?

Does she sometimes ask strangers embarrassing questions, without realising that they are embarrassing?

Does she respond to indirect hints to do something, for example to stop talking if the conversation is boring for the other person?

Can she fill in important missing information if someone else doesn't understand what is going on e.g. explaining the background of an event to a teacher?

Does she tell white lies? For example, will she say something contrary to what she thinks or knows in order to spare someone's feelings?

Can she express her ideas in more than one way, for example, if someone doesn't understand, can she explain things differently?

When talking to another person, does she seem to consider the other person's interests when deciding what to talk about, or does she pursue her own interests only?

Can she take turns appropriately in a conversation with another person, for example not constantly interrupting or talking over the other person?

Does she engage in flexible small talk, responding to what the other person says rather than having a fixed routine?

Can she behave appropriately with different groups of people, for example, behaving differently with strangers than with familiar people, or with adults or peers?

Is she naïve around people? For example, does she realise when someone is only teasing or playing a joke on her?

Does she recognise when she has hurt someone else's feelings?

### **Executive function scale**

Is she able to change task in work or play without being prompted to do so?

Does she express realistic plans or ideas (e.g. future goals)?

Does she like to have her day organised into set activities?

Does she find it difficult to keep her mind on something, and is easily distracted?

Is she good at keeping track of things she needs to do, for example when working on her own?

Does she get easily distracted in the middle of a task or game?

Does she have problems following verbal instructions, even though she is paying attention, unless they are kept simple and straightforward?

Does she think ahead about what she will need to do activities, rather than leaving things until the very last minute?

Does she often do something silly, just out of habit (e.g. going to fetch something from where it used to be, or going to an old classroom)?

Does she find it hard to stop saying or doing something once she has started?

Does she tend to be restless, and have problems keeping still for any length of time?

Does she collect together all work or play items before starting a task/game/project?

Does she act on impulse? For example, act without thinking, and do the first thing that comes to mind?

Is she able to work out sums in her head, for example, can she work out if she has enough money to buy several things

Does she talk on about things that interest her despite explicit attempts to change the topic of conversation?

### **General sociability scale**

Does she recognise when other people are happy or sad?

Does she use appropriate table manners?

Will she share things when asked, for example a textbook, or scissors and glue?

Does she initiate conversations with other people?

Does she have a best friend of the same sex?

## **APPENDIX III TOM TRAINING PROGRAMME**

### **Training sessions**

#### **Stage 1**

Before the child arrived, the researcher took photos of the room in which the training was taking place.

#### **Camera analogy**

The child was then shown the Polaroid camera, and the researcher took a photo. The child and researcher then watched the photo develop together. The researcher then pointed the camera at different locations in the room and asked the child to choose (from a pre-prepared range of photos) what the photo would show if it was taken.

#### **Introduction of doll**

The child was then introduced to Sarah the doll, and told that when Sarah looks at something, she has a thought in her head, like a picture of what she sees. Sarah then 'looked' at different locations in the room, and the child was asked to choose the correct picture from a pre-prepared range. The pictures were slotted into Sarah's head. This was repeated until the child got three consecutive trials correct.

#### **Transfer to real people**

The child was then told that it is the same with real people, but with real people we can't see the pictures as they are inside the person's head. These pictures are called thoughts. The researcher then exaggeratedly looked around the room, and asked the child to choose the photo which showed the correct 'thought-picture' which the researcher would have in their head given the direction they were looking in. Again, the child had to get three consecutive trials correct to move on to Stage 2.

The phrase written on the reminder card has some variation due to the input of the child, but typically said something like 'When people look at things, they get thoughts in their head like photos.'

#### **Stage 2**

#### **Review of last session**

The child read their reminder card, or the researcher read it to them, if they weren't comfortable with reading. Sarah was reintroduced, and the child was asked to choose the correct thought-pictures when Sarah looked at various objects.

### Thought pictures can stay in people's heads

Sarah then 'looked' at various objects, the child chose the correct thought picture, and Sarah then left the scene. The child was told that the picture in Sarah's head could stay there even though she wasn't looking anymore, and was asked what was on Sarah's thought picture. Sarah then returned, and the child was shown the photo, with feedback if wrong. After three consecutive correct trials, the child was allowed to move on.

Thought pictures can be used to look for things

An object (e.g. a small toy) was placed in a box, and the child chose a thought-picture to show where Sarah thought the object was. Sarah then left the scene, and the child was asked to describe her thought picture. Sarah then returned, and the child was told Sarah would use her thought-picture to look for the object. The child was then asked where Sarah would look for the object. They had to get three consecutive trials correct to move on to the next stage.

### Stories

The child did not have to pass this stage in order to move on to the next stage. This section was designed to consolidate the ideas involved and improve the likelihood of generalisation beyond the dolls. The child was told two illustrated stories about children. An example story is given.

*This is Kate. Kate has been reading her book. She puts it on the table while she goes to have supper. Which thought picture of her book does she have in her head? (child chose from four line drawings). Now Kate goes to have supper. After supper, she wants to read her book. Where does her thought picture show her book? Where will she look for her book? Where does she think her book is?*

*Let's look at the picture to see if you're right. She thinks what is on her thought picture.*

### False Belief

A standard unexpected transfer FB task was included at this point to assess spontaneous use of the photo analogy. An Elmo puppet, two boxes and a marble were used. Two children passed this task, one child in the MLD group and one in the ASD group.

The researcher and child then agreed together what should be written on the reminder card. This typically said something like 'People can use their thought-picture to look for things'.

### **Stage 3**

#### **Review of previous week**

Read reminder card, asked brief recap questions.

#### **Different people can have different thought-pictures in their heads**

Another doll – ‘Jo’ - was introduced, also with a slot in her head. The child was asked to choose thought-pictures for both dolls, as they looked in different directions. They had to get 3 consecutive trials correct.

#### **Perspective taking with photos**

The child was told ‘People take pictures in their heads of what they see. They can’t take a picture in their head of things they haven’t seen’. The purple cylinder was placed next to the toy crocodile, so that from one angle it obscured it. The child was asked to move around the room, and to choose the photos (from a pre-prepared range) that showed what could be seen from different angles. If they couldn’t do this, they were allowed to use the Polaroid to take photos and they watched the photos develop. They had to choose the right picture on three consecutive trials to move on.

#### **Thought-picture perspective taking**

The child was told ‘People are the same. Their thought picture will show what they can see. If they can’t see the toy, the toy won’t be on their thought-picture’. Different toys were placed beside the cylinder, and the dolls were placed at opposite sides. The child was asked initially to go and stand by the dolls in order to select the correct thought-picture, and later was asked to do it without moving. Line drawings were used as well as the photos. They had to get three consecutive trials correct in order to pass the stage.

#### **Stories**

*The child did not have to pass this section in order to move on, but responses were recorded.*

The child was shown a picture of two people looking at different objects and was asked what thought-pictures they had in their heads. No pictures were provided for them to choose from.

The reminder cards for this stage typically said ‘Different people can have different thought-pictures in their heads’.

### **Stage 4**

Before the child came in, the researcher prepared pictures of the toys in different locations in the room.

### Review of previous session

Read reminder card, asked recap questions.

### Thought pictures can be out of date

This was introduced with the camera. The researcher took a photo of a toy in a box, and then placed the photo face down on the table. The toy was then moved, and the child was asked where the toy would be on the photo. The child and researcher then looked at the photo together, and the researcher said ‘Even though we’ve moved the toy, when we took the photo the toy was in the box, so the photo still shows the toy in the box. The photo is out of date now. It doesn’t show how things really are. Well, it’s the same with thought pictures in people’s heads. Sometimes they can be out of date’.

A toy was then placed on the table, and the child was asked to choose a thought-picture for Sarah. The picture was slotted into Sarah’s head. Sarah then left the scene, and the child and researcher moved the toy. The child was then asked ‘Did Sarah see us move the toy? What does Sarah’s thought picture show?’ This was repeated until the child had three consecutive trials correct.

### Out of date thought pictures make people look in the wrong places

The scenario above was repeated, but this time when Sarah returned, she looked at her thought-picture and looked in the wrong place. After a demonstration, this was repeated and the child was asked what was on Sarah’s thought picture, and where Sarah would look for the toy. This was repeated with various toys and locations around the room until the child answered three consecutive trials correctly. This was the criterion to finish this stage.

### Stories

The child did not need to pass these to move onto the final stage. There were two illustrated unexpected transfer FB stories, with line drawing ‘thought-pictures’.

Reminder cards for this stage typically said ‘Sometimes thought-pictures are out-of-date. Then people look in the wrong places for things’.

### **Stage 5**

### Review of previous session

Read reminder card, recap of previous session.

### Thoughts are like pictures in your head

Scenarios like those used in stage 4 were repeated, but this time the child was told ‘The picture in her head shows what Sarah thinks.’ and after each scenario they were



asked 'So where does Sarah think the object is?'. Thought-pictures were still used throughout. They had to get three consecutive answers correct in order to finish.

#### Extension

If the child had passed all the previous stages with no difficulties an extension was introduced. This involved the idea that thought-pictures could be generated by other methods than looking – e.g. talking or reading. A doll was blindfolded, and was told where a toy was. The child then had to choose the correct thought picture. They had to get three consecutive answers correct to successfully complete the extension. Four children with MLD and 4 children with ASD passed the extension.

#### Final discussion

The final discussion consisted of making it clear that real people also had thoughts in their heads, and of finishing the reminder card, which was given to the child to keep. A 'thought-picture' of the child was stuck on the back of the card.

Toys used for training



Dolls (with thought-pictures in their heads) used for training



## **APPENDIX IV EF TRAINING PROGRAMME**

### **Training sessions**

#### **Stage 1**

##### **Machine analogy**

This stage introduced a scale model CAT truck, which had several removable tools. The child was told that the truck could do many things because it has different tools that it could use, and that the tool could be changed to make it do something different. Some stones, twigs and a plastic car were used. The child was asked to make the truck pick up the different categories of objects, choosing the correct tool each time and changing it on the truck. The various tools would only pick up one of the three objects – i.e. the stones could not be picked up with the implement which lifted the car, or vice versa. They had to indicate the correct tool on three consecutive trials to progress to the next section.

##### **Brain tools analogy**

The child was then told ‘People are like CAT trucks. We can do lots of things and sometimes we need to change how we do something, just like the truck. Apart from our arms and legs, we can also use our brain to do things. That’s our most flexible tool. People can change their tools when they do brain work, it’s just that we can’t see it’.

They were asked to select brain tools from a pre-prepared range to show what their brains enabled them to do in different situations, for example in P.E. (running around, playing games), in English lessons (writing, reading), at lunch time (eating) and at night (sleeping).

##### **Introduction of sorting**

The stones, twigs and car were then mixed up, and the truck was made to sort them out, by using various tools. A ‘brain card’ was then introduced. This was a cross-section of a head, with the brain visible. The brain tools were stuck onto the brain with blu-tack. The child was then told ‘We can sort things out using our brains too. When we sort out the cards by looking at what colour they are, we use our ‘Colour’ brain tool.’ The child was given a set of cards, and asked to use their colour tool to

sort them out. They were initially given lots of help, and were given different sets of cards until they had sorted 10 cards correctly.

They were then told 'Now we're going to change the brain tool to shape. Can you do it? Sometimes it's very hard to change. We'll take the *Colour* brain tool off, and change it to *Shape*, to remind us. We'll go 'all change' to help us remember.' An 'all change' procedure was demonstrated, where the researcher put both fingers on her temples and said 'all change'. The child was encouraged to repeat this each time they change the tool. Some children spontaneously came up with their own changing procedure, which was then used instead.

The change between two dimensions was repeated with various cards, until the child got at least 14/20 right. This was the criterion for moving onto Stage 2.

The reminder card was then completed by the experimenter in consultation with the child. For stage one it typically said 'People can do lots of different things using different brain tools'.

## **Stage 2**

Review of last stage

The reminder card was read by the child, or to the child, and a brief recap of the brain tool concept was carried out, with the child doing a simple change of brain tool.

Sometimes a brain tool gets stuck

The aim of this was to try and enable the child to experience the difficulty involved in changing a habitual way of doing something.

The truck was used to demonstrate the idea that a tool could get stuck. The child was told 'Sometimes a tool has been in place for a long time and it gets stuck. It's like the tool gets rusted on. The truck has been moving stones for a long time and now it has to pick up this car. Which tool does it need?' (child chooses tool). 'But it's really hard to take off the old tool because it's got stuck on after such a long time. We have to pull hard to get it off.'

The child was then shown the brain card again, and told that brain tools also get stuck sometimes and it's hard to change them. They were reminded of the 'all change' procedure as a way to help themselves change brain tools.

They then sorted 3 or 4 sets of cards by *Colour*. After doing this correctly, they were asked to change the tool to an *Animal* one, and the researcher said 'Is the old tool a bit stuck. Let's change it together'. If they failed to get 14/20 correct straight away,

this was repeated with different tools. If they passed, they were able to progress to the next stage.

### **Stories**

The child was then told 3 illustrated stories about brain tools getting stuck and being changed in real life. These stories were about a boy who has to get up early during the week, but can lie in at the weekends, another boy who goes to his maths classroom instead of a school trip, because usually he goes to maths first thing every day, and a girl who has to change the way she will get to school to prevent herself being late. The child was asked questions about the stories, but these were not assessed, although the responses were noted. The child was then asked if they themselves ever got brain tools stuck.

Reminder cards were then completed. They typically said 'Sometimes brain tools get stuck and it's hard to change them'.

### **Stage 3**

Review of last stage

Read phrase on reminder card and briefly recapped last session.

#### **STOP-CHANGE-GO**

Traffic light signals were introduced to help the child change brain tools. When the red STOP sign was held up, the child stopped, the orange CHANGE one meant they had to change their brain tool (using the all-change procedure) and the green was the signal to go. This was initially introduced with the truck.

Some brain tools are easier to use than others

The aim of this stage was to enable the children to experience the difference between using easier and harder 'brain tools' for themselves.

The child was asked to use their *Say what you See* tool, and was shown two cards repeatedly with pictures of a Sun and a Moon. They said 'Sun' on presentation of the sun, and 'Moon' on presentation of the moon. They were then asked (using the Stop-Change-Go sequence) to change to the *Say the Opposite* tool. They were told 'This is a hard tool to use. You can go slowly to help you'. They had to get 7/10 correct after the switch to progress to the next stage. This was repeated with girl/boy and knife/fork cards if they had difficulties.

If they completed this without difficulties, they were encouraged to change without using the brain tool cards.

They were then asked to use a *Do as I Say* brain tool, and to stand up or sit down on command. After changing brain tool they had to use a *Do the Opposite* tool, and sit down when the researcher said stand up and vice versa. They were allowed to correct themselves, and had to get four in a row right after the change to progress to the next level. This was repeated with different actions if they had problems. The child was then asked if they could think of any brain tools that were hard to use, or that were very easy to use.

The reminder card was then completed, typically it said 'Some brain tools are easy to use and some are harder. We can use Stop-Change-Go to help us change'.

#### **Stage 4**

Review of last stage

Reminder card and recap of last session.

Changing before finishing

This was demonstrated with the truck. The child was asked to move a pile of stones with the truck, but before they had finished they were told that there had been an emergency, a tree had fallen across the road and the truck needed to change and move the tree. The Stop-Change-Go was used to help them to change.

They were then told 'We can do just the same thing with our brain tools'. They sorted out a set of cards by the animal on them, and half way through were stopped (with the 'Stop' sign) and asked to change to the number brain tool. This was repeated with different sets of cards until the child got 5 in a row correct immediately after a change. This was the criterion for progression to the final stage.

Stories

The child was told 2 stories about children who had to change what they were doing before finishing. An example story is given:

*James was doing some writing at school. He was really enjoying writing a story.*

*Which tool was he using?*

*Before James had finished, the bell rang. It was time for maths. James had to change his tool even though he hadn't finished. Which tool does he need?*

*James found it hard to change because he was in the middle of writing his story, but he had to do maths instead, so he changed his tool. He went 'all change' and went to do his maths.*

The reminder card was then completed. It typically said 'Sometimes we have to change our brain tools before we finish doing something'.

## **Stage 5**

People have to decide for themselves which brain tools they will use

This stage was included to encourage the children to generate some brain tools themselves, and to try to generalise to more real life settings.

Initially the children were given familiar sets of cards (which had been used in earlier stages of the training) and asked what brain tools they could use to sort the cards out.

A blank set of brain tools were used, and the child or the researcher wrote/drew on them. To pass this stage, the child had to generate at least 3 different brain tools, 2 of which had to be for the same set of cards. If they had difficulties with this, they were told, 'Remember, we need different tools to do different jobs. If the truck had only one tool, it could only move sand, it couldn't move logs or cars. With the tool you've made, we could sort the cards by \_\_\_\_\_. What else could we sort them by? If I say 'all change', what tool could you change to?'

### **Extension**

If the child completed the above section without difficulty, a novel set of cards was introduced (which varied along at least 3 dimensions) and the child was asked to generate brain tools to sort those cards. All the children except one (who had ASD) were able to progress to this stage.

### **Stories.**

The child was told, 'People have to do this all the time. They have to decide all the different ways they could do something and then decide which to do. Sometimes they have to change.' They were told a story about a girl who had to think about all the ways she could get to her friend's house. They were then given a picture and asked to generate brain tools about it. For example:

*Bob wants to cross a lake to go to his house on the other side. He has different ways he could do it. What tools might he use? What would he do if it's a rainy day? What should he do if it was winter?*

Their responses were noted down.

Finally they were asked about themselves, and what brain tools they could use when they get home after school. These responses were also noted down.

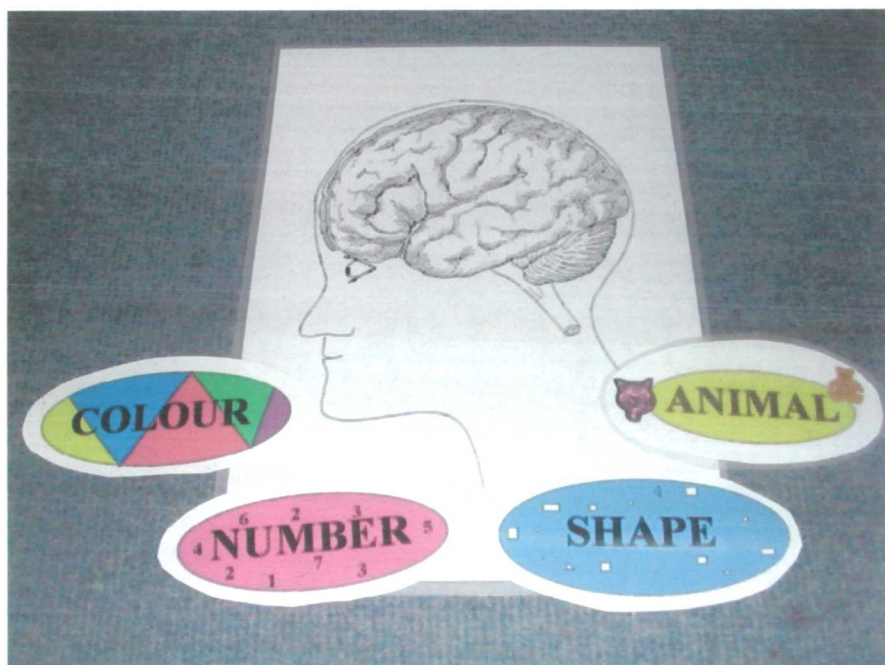
The reminder card was then finished, a typical phrase for this stage was 'Sometimes we have to decide for ourselves what brain tool we will use'. A 'Stop-Change-Go' picture was then stuck on the other side of the card, and it was given to the child to keep.



### Toy truck with removable tools



### Brain Tool card with brain tools





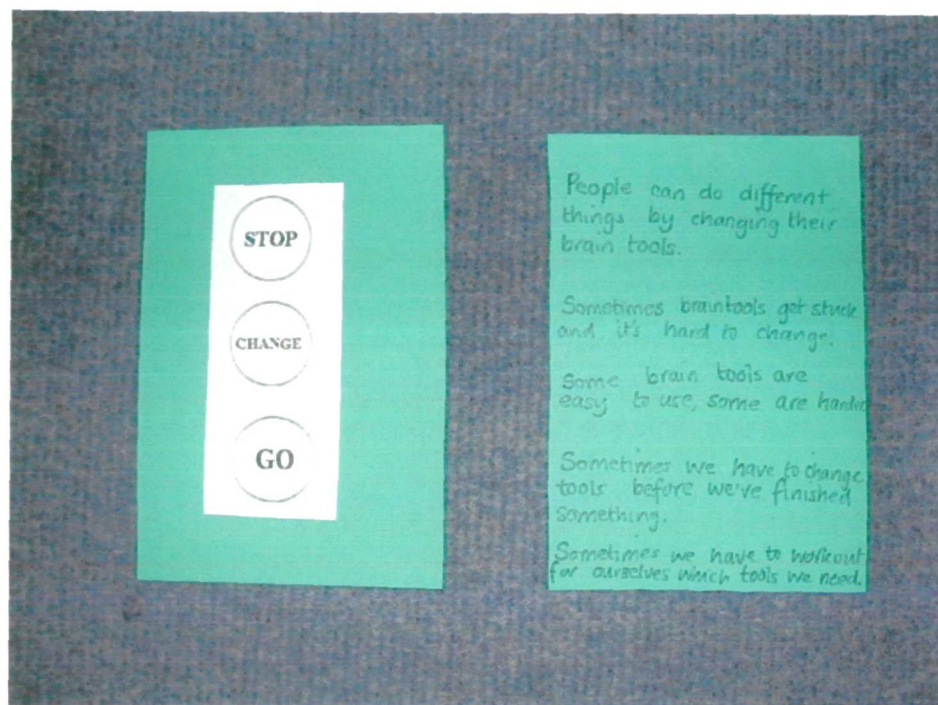
### Stop-Change-Go cards



### Example sorting cards



### Example reminder card, front and back



## **APPENDIX V NUMBER OF CHILDREN IN EACH TRAINED GROUP PASSING EACH TOM TASK AT EACH TIMEPOINT**

**Table V.I Number of children in the ASD group passing each task at each timepoint**

Task	ToM (N=10)			EF (N=10)			Controls (N=7)		
	Pre-test	Post-test	Follow-up	Pre-test	Post-test	Follow-up	Pre-test	Post-test	Follow-up
Unexpected transfer	2	9	8	3	4	3	1	3	2
Deceptive box-other	2	3	3	1	2	4	1	4	1
Deceptive box-self	5	6	3	3	4	7	3	1	1
Penny-Hiding	1	6	N/A	0	5	N/A	1	3	N/A
Seeing-Knowing	8	N/A	9	6	N/A	10	5	N/A	5
Know-guess-other	4	N/A	6	4	N/A	6	3	N/A	2
Know-guess-self	2	N/A	3	4	N/A	1	2	N/A	2

**Table V.II Number of children in the MLD group passing each task at each timepoint**

Task	ToM (N=6)			EF (N=7)			Controls (N=8)		
	Pre-test	Post-test	Follow-up	Pre-test	Post-test	Follow-up	Pre-test	Post-test	Follow-up
Unexpected transfer	0	5	4	1	2	2	0	6	5
Deceptive box-other	2	6	5	2	4	3	5	7	6
Deceptive box-self	4	6	4	3	5	5	7	6	7
Penny-Hiding	0	3	N/A	2	3	N/A	1	6	N/A
Seeing-Knowing	6	N/A	4	6	N/A	7	7	N/A	8
Know-guess-other	5	N/A	5	4	N/A	6	7	N/A	6
Know-guess-self	3	N/A	3	3	N/A	5	6	N/A	6

## **APPENDIX VI ADVANCED THEORY OF MIND TASKS**

### **Second order false belief task – the Chocolate Story**

This is Mary and her brother John.

Their Grandad has given them some chocolate to share.

“Put it away now,” says Grandad “You can have it when Mum says so”

John and Mary go inside and put the chocolate in the fridge.

Then they go out to play in the garden.

Later, John comes in for a glass of water.

He goes to the fridge and sees the chocolate.

He wants to keep the chocolate all for himself!

So he takes the chocolate out of the fridge and puts it in his bag.

**1<sup>st</sup> order FB: Where does Mary think the chocolate is?**

**Reality control: Where has John put the chocolate really?**

But, oh look, Mary is playing by the window now and she can see everything that John is doing!

She sees him put the chocolate in his bag!

John is so busy hiding the chocolate that he doesn't see Mary watching him through the window!

Later, Mum comes to call John and Mary in for tea.

She says they can have some of the chocolate now.

John and Mary come running into the kitchen.

**2<sup>nd</sup> order FB: Where does John think Mary will look for the chocolate?**

**Justification: Why does John think that?**

**Reality control: Did John see Mary watching him through the window?**

**Reality control: Where is the chocolate really?**

**Memory check: Where was the chocolate first of all?**

## **Examples of the Children's Strange Stories**

### **Theory of mind story**

Simon is a big liar. Simon's brother Jim knows this, he knows that Simon never tells the truth! Now yesterday Simon stole Jim's ping-pong paddle, and Jim knows Simon has hidden it somewhere, though he can't find it. He's very cross. So he finds Simon and he says, "Where is my ping-pong paddle? You must have hidden it either in the cupboard or under your bed, because I've looked everywhere else. Where is it, in the cupboard or under your bed?" Simon tells him the paddle is under his bed. Jim goes straight to the cupboard.

Question: Why does Jim look there?

### **Physical story**

A burglar is about to break into a jewellers' shop. He skilfully picks the lock on the shop door. Carefully he crawls under the electronic detector beam. If he breaks this beam it will set off the alarm. Quietly he opens the door of the store-room and sees the gems glittering. As he reaches out, however, he steps on something soft. He hears a screech and something small and furry runs out past him, towards the shop door. Immediately the alarm sounds.

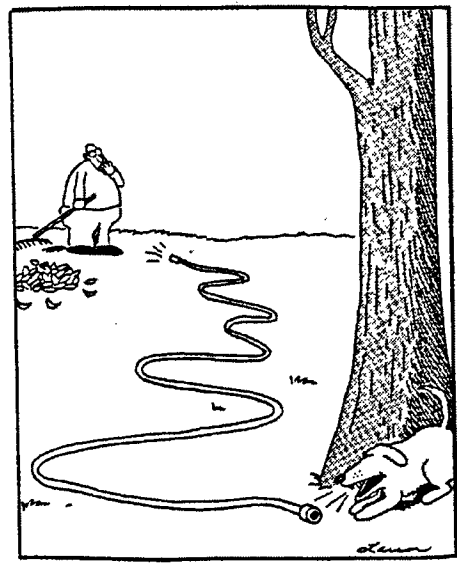
Question: Why did the alarm sound?

## Example children's cartoons

### Physical



### Theory of mind



## **Unexpected Stories**

The target sentences are in italics. Half the children were presented with one emotion for each story, whilst the other half were presented with the other emotion.

### **Positive expectation, negative reality**

#### **Expected**

Jane was waiting to go out to the cinema with her dad. They were going to see the new Pokemon film. They were going to go at half past three. Then her dad came in and said that he was sorry, but Jane had to go to the dentist's instead. Her dad told her to get her coat and put her shoes on. As Jane put her shoes on, she thought about the film and watched her dad getting ready to go. *She felt happy OR She felt sad.* They left the house together.

#### **Expected**

Lucy was going to go to a theme park. She had been looking forward to it for ages. She was going with her whole family and her friend Sandra. They were going to set off early on Saturday. On Friday Lucy and Sandra saw on TV that there were floods at the theme park. It would be closed all weekend. On Saturday Lucy woke up early. She thought about the theme park. *She felt happy OR She felt sad.* Then she got out of bed.

#### **Unexpected**

Jack was going to stay the night with Chris after school. They had been planning it for a long time. Jack had often visited Chris's house before and they always had fun. This was the first time he was actually going to stay the night. Outside school, Jack met Chris's mum. She didn't have time to tell Jack that Chris's cousins were coming to stay and so Jack couldn't come round today. As Jack went into his classroom, he saw Chris sitting at the table. *He felt happy OR He felt sad.* He went over to sit down.

#### **Unexpected**

Anne had entered a special competition. She had designed a poster telling people to recycle their rubbish. She had worked on her poster for a long time and it looked really good. Today in assembly they were going to announce the results. Before school, Anne met her teacher. She decided not to tell Anne that a girl in Class 8 had won the competition. On her way to assembly Anne thought about her poster and the competition. *She felt happy OR She felt sad.* She went into assembly.

### **Negative expectation, positive reality**

#### **Expected**

Simon was having a hard time at his new school. He was finding the work difficult and he didn't have many friends. Yesterday, his teacher Mrs. Green had told him off because he had been very slow. This morning, the headteacher asked him to come and see her after lunch. Later, Mrs. Green said that the headteacher had seen a poem that Simon had written. She had been very impressed with it. During lunch, Simon thought about the headteacher and school. *He felt happy OR He felt sad.* Then he went to her office.

#### **Expected**

Steven had a very important test to do at school. It was a maths test, and he had always found maths difficult. He had been studying for a long time the night before. He still wasn't sure he would know how to do all the problems. Maths was the first lesson in the morning. In the corridor, he met his friend Ben, who told him that the maths teacher was ill and so the test had been cancelled. As he entered the school, he thought about the maths test. *He felt happy OR He felt sad.* He went to the maths classroom.

#### **Unexpected**

Anita was going to the doctor. She had to have an injection to stop her getting ill. She didn't like having needles stuck into her. At the doctor's, Anita and her mother sat in the waiting room. Anita played with the toys and her mother read a magazine. No one explained to Anita that there wouldn't be a needle. Instead, Anita would be given the medicine on a sugar lump. Anita thought about going in to see the doctor. *She felt happy OR She felt sad.* Then the doctor called her name.

#### **Unexpected**

Neil had to go to the supermarket with his dad. It was Saturday morning. He really didn't like going to the supermarket. It always took a long time and they never bought anything interesting. His dad hadn't told Neil yet that his friend Joe had invited him round. He could spend the morning there instead of shopping. Neil thought about his Saturday morning. *He felt happy OR He felt sad.* Then he went out to the car with his dad.